Deciphering Genetic Susceptibility to Tuberculous Meningitis

By

Nicholas Bowker

Thesis presented in partial fulfilment of the requirements for the degree Master of Science in Human Genetics in the Faculty of Medicine and Health Sciences at the University of Stellenbosch

Supervisor: Dr. Marlo Möller
Faculty of Medicine and Health Sciences
Department of Molecular Biology and Human Genetics

Co-supervisors: Prof. Eileen Hoal, Dr. Craig Kinnear and Dr. Muneeb Salie
Faculty of Medicine and Health Sciences
Department of Molecular Biology and Human Genetics

December 2016

Declaration

By submitting this thesis/dissertation, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Signature: ……..

Date: …….

Table of Contents

Abstract

Tuberculous meningitis (TBM) is a type of extrapulmonary tuberculosis (TB) which leads to inflammation of the meninges through small lesions called Rich foci. TBM represents 1% of total TB disease, with the age of onset being around 2-5 years of age. Disease development remains poorly understood, including the mechanism of dissemination across the blood-brain barrier.

This study concentrated on the involvement of the host genome in TBM susceptibility. We hypothesised that multiple common variants of moderate effect size are more likely to influence TBM susceptibility than rare variants of large effect. The SAC are a 5-way admixed population with genetic contributions from European, East Asian, South Asian, Bantu-speaking African and KhoeSan groups.

Two technologies were used to genotype single nucleotide polymorphisms (SNPs) of varying frequency: exome sequencing and the Illumina® Multi-Ethnic Genotyping Array (MEGA). Exome sequencing involves capture of only the protein-coding regions of the human genome, constituting approximately 1% of the genome. The depth of coverage of exome sequencing enabled the detection of coding SNPs of lower frequency (<1%) which were assessed for association with TBM susceptibility. Ten TBM cases and 10 healthy controls were exome sequenced, with all study participants being from the SAC population.

Gene set association tests SKAT-O and SKAT Common Rare were used to assess the association of rare SNPs and the cumulative effect of both common and rare SNPs with susceptibility to TBM, respectively. The SKAT-O analysis included 8 322 gene sets comprising 16 728 SNPs, which did not yield any associations with TBM susceptibility. Ingenuity Pathway Analysis (IPA) of the top-hits of the SKAT-O analysis showed that *NOD2* and *CYP4F2* are both important in TBM pathogenesis and highlighted these as targets for future study. The SKAT Common Rare analysis included 13 270 gene sets comprised of 53 239 SNPs and was *CCP110* associated ($p = 5.89x10^{-6}$) with TBM susceptibility. In addition, a number of top-hit genes ascribed to the development of the central nervous system (CNS) and innate immune system regulation were highlighted.

The role of common variants (>5%) in TBM susceptibility were assessed by conducting a genome-wide association study (GWAS). A total of 123 TBM cases, 400 pulmonary TB (pTB) cases and 477 healthy controls were genotyped on the MEGA array. A GWAS comparing 114 TBM cases

to 395 healthy controls showed no association with TBM susceptibility. A second analysis comparing 114 TBM cases to 382 pTB cases was conducted to investigate variants associated with different TB phenotypes. No significant associations were found with progression from pTB to TBM.

This study represents the first exome sequencing and GWAS of a TBM cohort and has identified a single previously undescribed association with TBM susceptibility. These results further our understanding of TBM in terms of both SNPs and genes that influence susceptibility. In addition, a number of candidate genes involved in innate immunity have been identified using IPA for further genotypic and functional investigation.

Opsomming

Tuberkuleuse meningitis (TBM) is 'n vorm van ekstra-pulmonêre tuberkulose (TB) wat lei tot inflammasie van die meninges deur middel van die vorming van klein letsels bekend as Rich fokusse. TBM verteenwoordig 1% van die totale TB epidemie, met ‘n aanvangsouderdom van ongeveer 2-5 jarige ouderdom. Die ontwikkeling van die siekte, asook die meganisme van beweging van bakterieë oor die bloed-breinskans, word huidiglik nie volkome verstaan nie.

Hierdie studie het die rol van die gasheer genoom in TBM vatbaarheid ondersoek. Die hipotese was dat daar ‘n groter waarskynlikheid is dat veelvuldige algemene variante met ‘n matige effekgrootte ‘n rol sal speel in TBM vatbaarheid as raar variant met ‘n groter effekgrootte. Die SAK het genetiese bydraes van 5 verskillende bevolkings, naamlik Europese, Oos-Asiatiese, Suid-Asiatiese, Bantu-sprekende Afrikane en KhoeSan groepe.

Twee tegnologieë was gebruik om enkel nukleotied polimorfismes (ENPs) van wisselende frekwensie te genotipeer: eksoom volgordebepaling en die Illumina® Multi-Ethnic Genotyping Array (MEGA). Eksoom volgordebepaling ondersoek slegs proteïen koderende areas van die menslike genoom, wat ongeveer 1% van die genoom uitmaak. Die diepte van dekking van die eksoom volgordebepaling het die opsporing van koderende ENPs met lae frekwensies (<1%) moontlik gemaak en hierdie variante se assosiasie met TBM vatbaarheid is geëvalueer. Eksoom volgordebepaling is uitgevoer op 10 TBM pasiënte en 10 gesonde kontroles van die SAK bevolking.

Die geen-stel assosiasie toetse SKAT-O en SKAT Common Rare was gebruik om die assosiasie van skaars ENPs sowel as die kumulatiewe effek van beide algemene en skaars ENPs met TBM vatbaarheid te evalueer. Die SKAT-O analise het 8 322 geen-stelle, bestaande uit 16 728 ENPs, ingesluit wat nie enige assosiasies met TBM vatbaarheid opgelewer het nie. Ingenuity Pathway Analysis (IPA) van die gene, aangedui deur die SKAT-O analise, het getoon dat beide *NOD2* en *CYP4F2* belangrike rolle in TBM siekte ontwikkeling speel en daarom kandidate vir toekomstige studies is. Die SKAT Common Rare analise het 13 270 geen-stelle, bestaande uit 53 239 ENPs, ingesluit en *CCP110* was geassosieer ($p = 5.89x10^{-6}$) met TBM vatbaarheid. Verder is 'n aantal gene betrokke by die ontwikkeling van die senuweestelsel en regulasie van die aangebore immuunsisteem uitgelig.

Die rol van algemene variante (> 5%) in TBM vatbaarheid was ondersoek deur 'n genoom-wye assosiasie studie (GWAS) uit te voer. 'n Totaal van 123 TBM gevalle, 400 pulmonêre TB (PTB) gevalle en 477 gesonde kontroles is genotipeer op die MEGA. Die GWAS analise, wat 114 TBM pasiënte met 395 gesonde kontroles vergelyk het, het nie enige assosiasie met TBM vatbaarheid gevind nie. 'n Tweede analise wat 114 TBM pasiënte met 382 PTB pasiënte vergelyk het was uitgevoer om assosiasie van variante met verskillende TB fenotipes te ondersoek. Geen assosiasie met progressie vanaf pTB tot TBM was gevind nie.

Die huidige studie verteenwoordig die eerste eksoom volgordebepaling en GWAS van ‘n TBM groep en het ‘n enkele voorheen onbeskryfde assosiasie met TBM vatbaarheid geïdentifiseer. Hierdie resultate bevorder ons begrip van TBM in terme van ENPs sowel as gene wat siektevatbaarheid beïnvloed. Daarbenewens was 'n aantal gene betrokke by aangebore immuniteit met behulp van IPA geïdentifiseer vir verdere genotipiese en funksionele ondersoeke.

Acknowledgements

My supervisor Dr. Marlo Möller for her constant availability and willingness to help me whenever I needed. Your commitment and drive for success really pushed me to do my best. Finally, a particular mention for the advice that you have given me concerning belief in my results and in myself has really stuck with me.

My co-supervisor Prof. Eileen Hoal van Helden and HOD Prof. Paul van Helden, for your support both academically and financially and willingness to edit the endless drafts of both my thesis and research articles. I am very grateful to the both of you for the opportunity that I was afforded to attend the Exome Sequencing course in Cambridge, UK.

To Dr. Brenna Henn for providing the KhoeSan reference population used in our admixture calculations. To the Hussman Institute for Human Genomics and Christian-Albrechts University of Kiel for running the genotyping array and exome sequencing experiments respectively. Also to Dr. Genevieve Wojcik for your invaluable assistance with the recalling of the MEGA array dataset.

The Harry Crossley Foundation, National Research Foundation, Stellenbosch University and Dr. Dr. Marlo Möller for the bursary contributions, without which I would not have been able to afford further education. The study individuals that consented to take part in this research and the clinical practitioners involved in recruitment and patient care: Dr. Ronald van Toorn, Dr. Reagan Solomons and Dr John Schoeman.

Special thanks to my friends that went through MSc. with me, our daily lunch trips really helped to relieve the stress and I don't think I have ever laughed so much in my life! You guys are fantastic and I have so many memories!

Finally to my parents, who have always supported and encouraged me in whatever I have done. Your support has gotten me further than I actually thought possible considering the challenges I have faced with dyslexia. To my girlfriend, Jessica, your daily support and encouragement have made my life so much easier. Thanks for putting up with all the stupidity and daftness that seems to crop up every day from me. Thanks for being the most supportive, kind and generous person in my life, you have no idea how much it means to me.

List of Figures

List of Tables

Glossary

A2M	…. Alpha-2-Macroglobulin
ADA	…. Adenosine Deaminase
ADAMTS2	…. ADAM metalloprotease with thrombospondin type 1 motif
ANKRD62	…. Ankyrin repeat domain 62
ATG10	…. Autophagy-related 10
B. bergdorferi	…. *Borrelia bergdorferi*
BBB	…. Blood-Brain Barrier
BCB	…. Blood Cerebrospinal Fluid Barrier
BCG	…. Bacille Calmette-Guérin
Bp	…. Base-pairs
BWA	…. Burrows Wheeler Aligner
C4bC2a	…. C3 convertase enzyme complex
C4BPA	…. Complement component 4 binding protein, alpha
Ca^{2+}	…. Calcium
CAAPA	…. Consortium on Asthma among African-ancestry Populations in the Americas
CAST	…. Cohort Allelic Sums Test
CBX3	…. Chromobox-3
CCL2	…. C-C Motif Chemokine Ligand 2
CCP110	…. Centriolar Coiled-Coil Protein 110kDa
CCR5	…. C-C Chemokine Receptor 5
$CD4^{+}$	…. Cluster of Differentiation 4^{+}

CDC	…. Centres for Disease Control and Prevention
CDH1	…. Cadherin-1
CDH5	…. Vascular Endothelial Cadherin
CEU	…. Caucasian Europeans from Utah, USA
CFH	…. Complement Factor H
CFHR3	…. Complement Factor H Receptor 3
CHB	…. Han Chinese from Beijing, China
CI	…. Confidence Interval
CNS	…. Central Nervous System
CRP	…. C - reactive protein
CSF	…. Cerebrospinal Fluid
CT	…. Computer Tomography
CX3CL1	…. Fractalkine receptor
CX3CR1	…. Chemokine (C-X3-C motif) receptor 1
CXCL9	…. Chemokine C-X-C Motif Ligand 9
CYFIP1	…. Cytoplasmic FMR1 Interacting Protein 1
CYP4F2	…. Cytochrome P450 Family 4 Subfamily F Member 2
CYP4F3	…. Cytochrome P450 Family 4 Subfamily F Member 3
DAP12	…. DNAX-activating protein 12kDa
DNA	…. Deoxyribonucleic Acid
DNMT1	…. DNA (cytosine-5) methyltransferase 1
EMMAX	…. Efficient mixed-model association expedited
EMT	…. Endothelial-Mesenchymal Transition

ERK 1/2	…. Extracellular Signal-Related Kinases 1/2
ESP	…. Exome Sequencing Project 6500
ExAC	…. Exome Aggregation Consortium
EZH2	…. Enhancer of Zeste 2 Polycomb Repressive Complex 2 subunit
FARP2	…. FERM, ARH/RhoGEF And Pleckstrin Domain Protein 2
FDR	…. False-Discovery Rate
FI	…. Serine protease factor I
FMR1	…. Fragile X Mental Retardation 1
FOXA1	…. Forkhead Box A1
FZD5	…. Frizzled-5
GBS	…. Group B *Streptococcus*
GIH	…. Gujarati Indians from Houston, Texas, USA
GNRH1	.... Gonadotropin Releasing Hormone 1
GWAS	…. Genome-Wide Association Study
H. influenzae	…. *Haemophilus influenzae*
HBHA	…. Heparin-Binding Haemagglutinin (Mycobacterial Protein)
HIHG	…. Hussman Institute for Human Genomics
HIV	…. Human Immunodeficiency Virus
HLA-DRB1	…. Human Leukocyte Antigen DR Beta 1 Chain
HWE	…. Hardy-Weinberg Equilibrium
IBD	…. Identity by descent
IFITM2	…. Interferon induced transmembrane protein 2
IFN-γ	…. Interferon-γ

IGRA	…. Interferon-γ Release Assay
IGV	…. Integrative Genomics Viewer
IL-	…. Interleukin
IPA	…. Ingenuity Pathway Analysis
IQR	…. Inter-quartile Range
IRAK3	…. Interleukin 1 receptor associated kinase 3 (A.K.A. *IRAK-M*)
IRIS	…. Immune Reconstitution Inflammatory Syndrome
ITGB2	…. Integrin beta-2 (also known as CD18 [Cluster of Differentiation 18])
Kb	…. Kilo-base
LD	…. Linkage Disequilibrium
LEF	…. Lymphoid Enhancer Factor
LINC01492	…. Long Intergenic Non-protein Coding RNA 1492
LoFTool	…. Loss-of-Function Tool
LPS	…. Lipopolysaccharide
LTA4H	…. Leukotriene A4 Hydrolase
LTB4	…. Leukotriene B4
LWK	…. Luhya from Webuye, Kenya
LXA_4	…. Lipoxin A4
LXB_4	…. Lipoxin B4
M. smegmatis	…. *Mycobacterium smegmatis*
M.bovis	…. *Mycobacterium bovis*
M.marinum	…. *Mycobacterium marinum*

M.tb	…. *Mycobacterium tuberculosis*
MACF1	…. Microtubule-Actin Crosslinking Factor 1
MAF	…. Minor Allele Frequency
MAPK	…. Mitogen Activated Protein Kinase
MARCO	…. Macrophage receptor with collagenous structure
MDR	…. Multi-Drug Resistant
MEGA	…. Multi-Ethnic Genotyping Array
MMP2	…. Matrix Metalloprotease-2
MMP9	…. Matrix Metalloprotease-9
MMP-9	…. Matrix metalloproteinase-9
MRI	…. Magnetic Resonance Imaging
MS	…. Multiple Sclerosis
MTAP	…. Methylthioadenosine Phosphorylase
MyD88	…. Myeloid Differentiation Primary Response 88
N. meningitidis	…. *Neisseria Meningitidis*
NAAT	…. Nucleic Acid Amplification Test
NFκB	…. Nuclear factor κ B subunit 1
NK	…. Natural Killer
NKIRAS1	…. NF-κB inhibitor interacting Ras-like 1
NLGN4X	…. Neuroligin 4 X-linked
NO	…. Nitric Oxide
NOD2	…. Nucleotide binding and oligomerisation domain 2 (AKA *CARD15*)

NOS-2	…. Nitric Oxide Synthase 2
NRAMP1	…. Natural Resistance-Associated Macrophage Protein 1
OCLN	…. Occludin
OR	…. Odds Ratio
ORF	…. Open Reading Frame
P2X7	…. Purinergic Receptor
PA	…. Propionic Acidemia
PAGE	…. Population Architecture in Genomics and Epidemiology
PAM2	…. PAM2Cys-SKKK Ligand
PAMP	…. Pathogen-associated molecular pattern
PCA	…. Principal Component Analysis
PCCA	…. Propionyl-CoA carboxylase alpha subunit
PCR	…. Polymerase Chain Reaction
PDCD4	…. Programmed cell death 4
PGL	…. Phenolic Glycolipid
PHC1	…. Polyhomeotic homolog-1
PKS 15/1	…. Mycobacterial Polyketide Synthase 15/1 gene
PLD	…. Phospholipase D
PorA	…. Porin-A
Prop	…. Proportion
pTB	…. Pulmonary Tuberculosis
QC	…. Quality Control
ROR1	…. Receptor tyrosine kinase-like orphan receptor 1

RtPCR	…. Real-time Polymerase Chain Reaction
Rv0960	…. vapC9 (Mycobacterial gene)
Rv1001	…. arcA (Mycobacterial gene)
S. pneumoniae	…. *Streptococcus pneumoniae*
SAC	…. South African Coloured
SAN	…. Khomani San from Upington, South Africa
SD	…. Standard Deviation
SERGEF	…. Secretion regulating guanine nucleotide exchange factor
SETDB1	…. Histone-lysine N methyltransferase
SKAT	…. Sequence Kernel Association Test
SNAI1	…. Snail-1
SNAI2	…. Snail transcription factor 2 (Slug)
SNP	…. Single Nucleotide Polymorphism
SRCR	…. Scavenger Receptor Cysteine-Rich domain
TB	…. Tuberculosis
TBM	…. Tuberculous Meningitis
TCF	…. T-cell Factor
TDM	…. Trehalose Dimycolate
TE	…. Tris-EDTA Buffer
TGF-β1	…. Transforming Growth Factor Beta 1
Th1	…. T-Helper 1 Lymphocytes
TIRAP	…. Toll-Interleukin 1 Receptor Adaptor Protein
TJP-1	…. Tight junction protein 1

TLR	…. Toll-like Receptor
TNF-α	…. Tumour Necrosis Factor-α
TST	…. Tuberculin Skin Test
USF1	…. Upstream Transcription Factor 1
VCF	…. Variant Call Format
VEGF	…. Vascular Endothelial Growth Factor
VEP	…. Variant Effect Predictor
WES	…. Whole Exome Sequencing
WHO	…. World Health Organisation
WNT5A	…. Wingless Homolog 5A
XDR	…. Extensively-Drug Resistant
Xpert MTB/RIF	…. *Mycobacterium tuberculosis*/Rifampicin Xpert Assay
ZEB1	…. Zinc Finger E-Box Binding Homeobox 1
ZFHX3	…. Zinc Finger Homeobox-3

General Introduction Chapter

1

Chapter 1: Contents

Chapter 1: Introduction

1.1 Global Epidemic

Tuberculosis (TB) ranks alongside human immunodeficiency virus (HIV) as the leading causes of death from infectious agents and remains a major global health threat [2]. In 1993, the World Health Organisation (WHO) declared TB a global public health emergency. An estimated 9.6 million TB cases and a further 1.5 million deaths due to TB were reported in 2014 [1]. By the year 2014, health care initiatives had improved, to varying degrees, in many countries, but global statistics remained virtually unaltered, with many developing countries still greatly affected by the disease as shown in Figure 1.1.

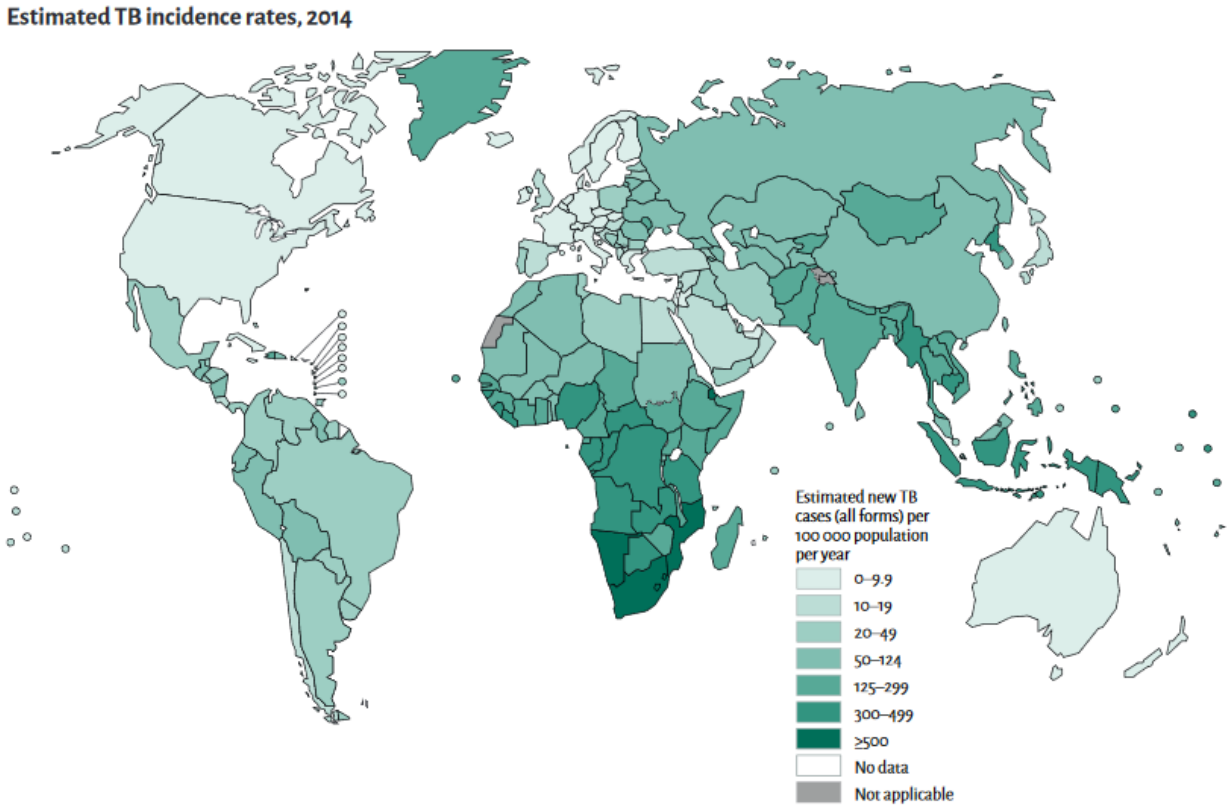

Figure 1.1: Estimated TB incidence rates worldwide in 2014. Mainly developing countries are affected by the disease, with developed countries still being affected but with a far decreased incidence. Adapted from [2].

The majority of TB incident cases were from developing countries with Asian and African countries having the highest incidences. The six countries with the highest number of first occurrence cases in 2014 were India, Indonesia, China, Nigeria, Pakistan and South Africa,

respectively [2]. The causative agent of TB disease is *Mycobacterium tuberculosis* (*M.tb*) and inquests into TB disease aetiology found a complex interaction between factors of the bacterium, human host and the environment [3, 4].

1.2 The impact of TB in South Africa

South Africa has one of the highest TB incidence and subsequent mortality rates in the world. This is in part due to a high prevalence of HIV nationwide (19.1% in 2013) [5]. The complication of HIV has hindered progress towards lowering the national TB incidence rate as immuno-compromised individuals are more susceptible to TB disease [6]. TB can be treated successfully, but despite South Africa's 78% treatment success rate, the death rate from the disease remains high, the population has poor adherence to treatment regimens and the spread of multi- (MDR) and extensively drug resistant (XDR) strains has placed a massive burden on the country's already over-taxed medical resources [7].

In 2014 the Western Cape Province had the fourth highest incidence rate in the country with 710 new cases per 100 000 [8]. The province with the highest incidence rate was KwaZulu-Natal Province with an incidence rate of 1076 per 100 000. The estimated TB incidence for the country in 2010 was 981 per 100 000 (Figure 1.2) with a prevalence of 795 per 100 000 [9].

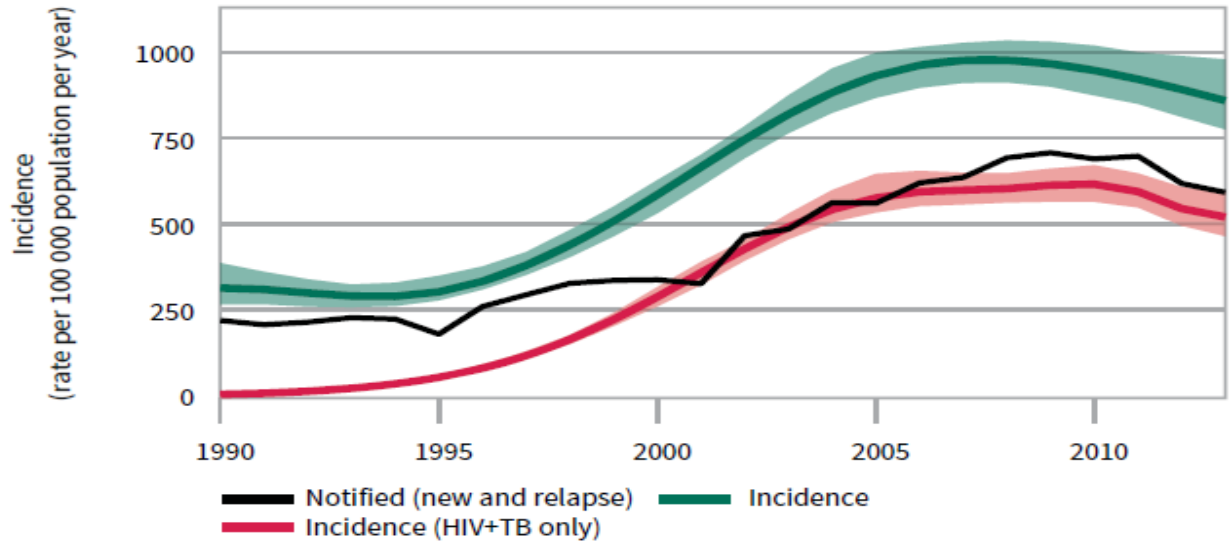

Figure 1.2: TB incidence rates (per 100 000) in South Africa showing the different proportions of TB incidence. Adapted from [9].

1.3 Tuberculous Meningitis

1.3.1 Background

Tuberculous meningitis (TBM) is a form of extrapulmonary TB leading to severe meningeal inflammation through small lesions called Rich foci [10]. The causative agent remains the same as for pulmonary TB (pTB), but the definitive process of dissemination has not yet been elucidated [11]. TBM infections represent a minimal proportion of total TB infection, approximately 1% [12]. Additionally, TBM constitutes only 5-15% of all reported extrapulmonary TB cases worldwide, yet still remains the form of TB that has the highest mortality rate at 20 - 25%. Many survivors suffer from high rates of neurological sequelae as a result of the disease [13, 14]. The approximate TBM incidence rate in the Western Cape was 31.5/100 000 in children under the age of 1 year in 1992 [15]. This figure has not changed, despite the common practice of neonatal vaccination using Bacille Calmette-Guérin (BCG) [16], leading health practitioners to question its efficacy in immunisation against the disease.

1.3.2 Pathogenesis

TBM is the most severe form of TB owing to its high mortality rate and is most frequently observed in children under the age of 5 years [11, 17]. TBM often develops within 3 months post initial infection and outcome is generally very poor, despite modern anti-tuberculosis drug regimens [16, 18]. Survivors of the disease frequently suffer from debilitating neurological sequelae such as cognitive disabilities, optic atrophy and motor system deficits [19].

TBM pathology commences with the inhalation of *M.tb*, which is subsequently ingested by alveolar macrophages, where immune system inflammation is initiated [19]. Following this, there is increased inflammation in the lung and subsequent dissemination to the lymphatic system where a short period of bacteraemia occurs which can lead to the haematogenous spread of the bacterial infection to the meninges and other organs [20, 21]. Haematogenous dissemination leads to the escape of the bacteria from blood vessels in the sub-arachnoid space, located in the meninges, and subsequent inflammation of the surrounding tissue. The meninges are a protective layer surrounding the brain and spinal cord which acts cooperatively with the cerebrospinal fluid (CSF) to protect against mechanical injury and maintain blood supply [12]. In the case of dissemination to the meninges, small sub-pia mater foci are formed. These foci are named Rich foci after the

original studies done by Rich and McCordick on guinea pigs and rabbits [21, 22]. Rich foci are caseous and granulomatous in nature and can be found in either the brain cortex, meninges or spinal cord [10, 20]. Rich foci most often follow the blood vessels along their length. Rich and McCordick defined TBM and miliary TB as two independent diseases, due to cases of miliary TB where TBM did not develop even though many other organs contained tubercles. Even if tubercules were found in the meningeal region, this would often not lead to the development of TBM, suggesting that miliary TB and TBM were independent diseases [22]. Other researchers believe that the two diseases share a pathophysiological relationship where the Rich focus remains a central factor, but that miliary TB is an exacerbated form of dissemination to other organs [19]. Additionally, the Rich focus mode of entry does not explain the frequent association of miliary TB and TBM [23]. Donald et al. (2005) concluded that miliary TB assists in TBM development in children and equally increases the chance of Rich focus development and resultant rupture, leading to TBM [20].

The second phase of TBM development involves the rupture of the Rich focus into the subarachnoid space causing meningitis which, if untreated, can lead to serious neurological damage [10]. The structure of the meninges and the sub-arachnoid space is depicted in Figure 1.3.

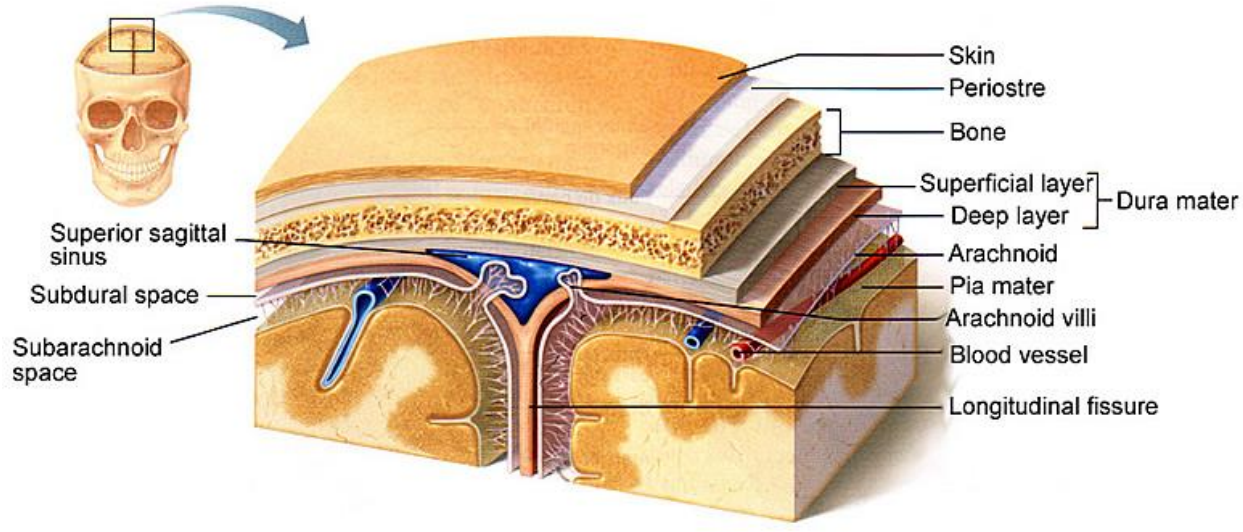

Figure 1.3: The layers of the meninges depicting the Arachnoid layer, pia mater and blood vessels that aid in the haematogenous dissemination of TBM. Adapted from [24]

In other forms of bacterial meningitis, a focal point such as the Rich focus is not observed; instead, direct haematogenous spread takes place [23].

Following the release of *M.tb* from the granulomatous Rich foci into the sub-arachnoid space, a dense exudate forms which can be observed on magnetic resonance imaging (MRI) brain scans along with abnormal meningeal enhancement also seen using imaging technologies. The exudate contains many lymphocytes, plasma cells and macrophages due to the high bacterial load in the area and develops around arteries, cranial nerves and the spinal cord [20, 21, 24]. CSF drainage is therefore impaired and outflow prevented, leading to increased intracranial pressure, hydrocephalus, ischemia and infarction [19]. Rich foci are not commonly found in the lower brain regions in close proximity to the exudate. Instead the exudate is thought to be located in the basal areas of the brain due to normal CSF flow patterns before blockage occurs [22]. Vasculitis can also occur of both large and small blood vessels leading to stroke-like symptoms in some patients and may also result in paralyses such as hemiplegia and paraplegia [21]. CSF flow restriction therefore results in the severe and debilitating neurological sequelae observed [19, 21, 23]. It is at this point when many neurological symptoms are often observed by the patient with disease progression too advanced and thus poor outcomes [25].

1.3.3 Bacterial Migration

One of the most contested points in TBM research is the mechanism of bacterial migration from the bloodstream across the blood-brain barrier (BBB). The BBB is composed of endothelial cells arranged in tight junctions surrounded by a basement membrane and astroglial end-feet [26]. The unique structure of the BBB allows for efficient isolation of the Central Nervous System (CNS) from the rest of the circulatory system (Figure 1.4) [27].

Through the study of other forms of meningitis three major mechanisms of migration across the barrier have been elucidated: transcellular, paracellular and ‘Trojan Horse’ migration [10]. Organisms such as *Haemophilus influenzae* use the transcellular method of migration which involves binding to the endothelial cells of the BBB through host laminin receptors via expression of phosphorylcholine to mimic platelet-activating factor, a molecule readily allowed to cross the BBB [28]. The Trojan Horse mechanism is a hallmark observed in fungal meningitis where traversal of the membrane depends on inositol, found in high concentrations in human and animal brains [29]. Considering the advanced inositol metabolism of mycobacteria, this represents a plausible method of traversal for mycobacterial species [10].

In vitro studies have shown that *M.tb* infected monocytes cross the walls of alveoli more efficiently than uninfected ones or single mycobacteria [30]. While the hypothesis of BBB traversal through infected monocytes may seem plausible, a major flaw is that during the onset of TBM such cellular traffic is heavily restricted across the BBB thereby limiting access to the CNS [27]. Early studies showed that free *M.tb* organisms can invade the CNS without the need for encapsulation and protection by monocytes [31]. More recent studies in murine models with Integrin beta-2 (*ITGB2*) [also known as CD18] knockouts corroborated these findings [32]. *M.tb* can infect host microglia which are localised in the CNS [33]. Microglia therefore perform a function similar to that of a macrophage as a host phagocytic cell in the CNS, suggesting that microglia provide a suitable micro-environment for the replication and maintenance of the mycobacterium.

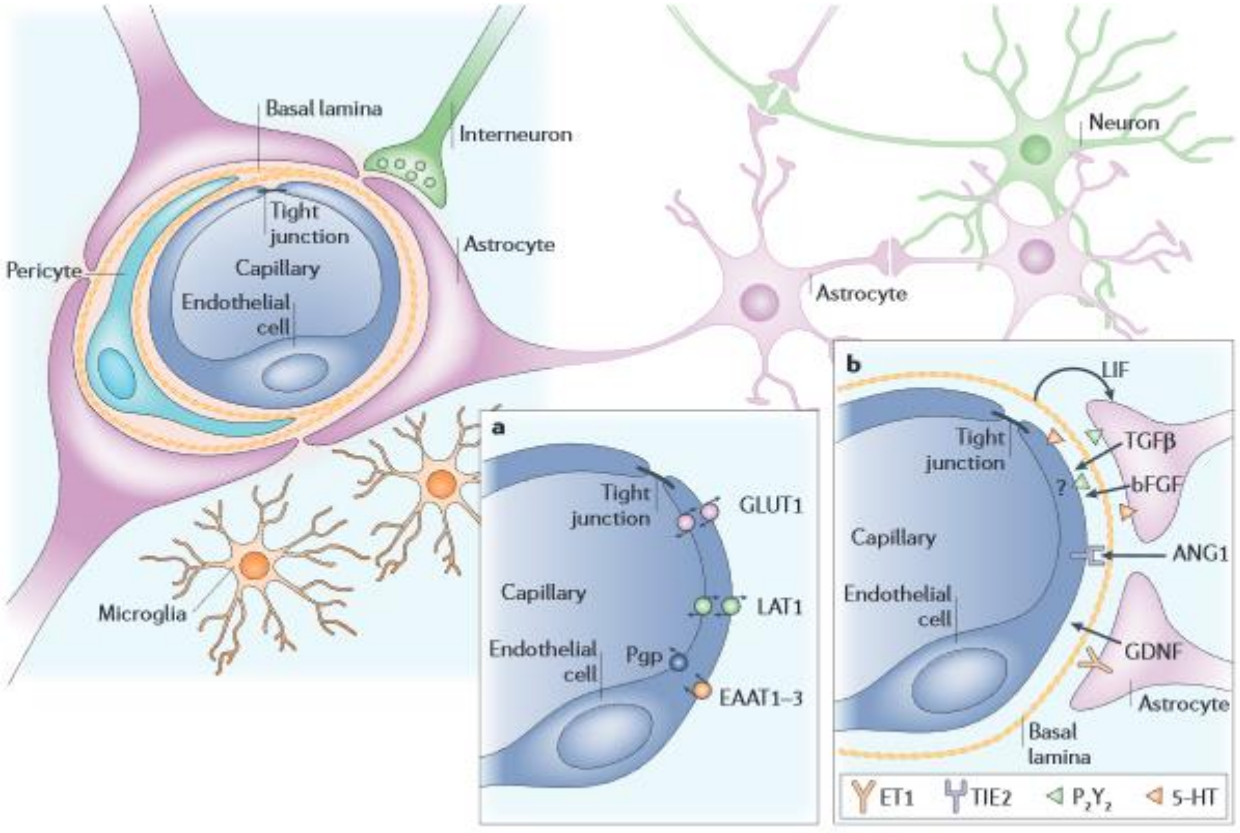

Figure 1.4: The structure of the BBB. (a.) Structure and transporters of the endothelial cells of the BBB. (b.) Depiction of the interaction between astroglial end-feet and endothelial cells needed for maintenance of the BBB. Adapted from [26].

Molecules upregulated during early TBM infection may play a role in TBM dissemination to the meninges and it was found that heparin-binding haemagglutinin (HBHA) adhesin is used by *M.tb*

to bind to the BBB's structurally integral epithelial cells using a C-terminal domain [34]. HBHA deficient strains have an ineffective ability to escape from the lungs, therefore suggesting a role for HBHA in the spread of *M.tb*, and for strain differences in disease severity [35, 36]. Several other molecules are used by *M.tb* in order to traverse the BBB, namely matrix metalloproteinase-9 (*MMP-9*) and vascular endothelial growth factor (*VEGF*) [37]. MMP-9 is commonly found in the basement membrane and functions in BBB degradation. It is not often present in the CSF but may be upregulated in conjunction with tumour necrosis factor-α (TNF-α) during early TBM development. VEGF is a multifunctional cytokine that is involved in inflammation, angiogenesis and stimulation of MMP-9 secretion. It was associated with meningitis, particularly in TBM patients (mean ± SD = 144.4 ± 75.1 pg/ml, $p < 0.01$) [38].

In vitro modelling of processes such as BBB invasion by bacteria enhance our understanding of the molecular mechanisms involved. In order to determine the method of invasion of brain microvascular endothelial cells, a major component of the BBB, an invasion assay was set up using a monolayer of primary endothelial cells incubated with both *M.tb* and the non-virulent strain *M. smegmatis* (Figure 1.5) [39]. *M.tb* invaded and traversed the BBB model with far greater efficiency than did *M. smegmatis*. This was achieved through host-cell actin cytoskeletal rearrangements hypothesised to occur through interaction between bacterial virulence factors and host recognition receptors. Gene-expression analysis found that 33 genes in a genetic island of the mycobacterial genome were upregulated, 18 of which belonged to a previously described island (*Rv0960-Rv1001)*. The BBB remains a crucial barrier to separating the brain from the rest of the circulatory system. Despite the fact that one can have TBM and not have BBB breach or damage, it still remains a physical hindrance and does help to reduce bacterial load.

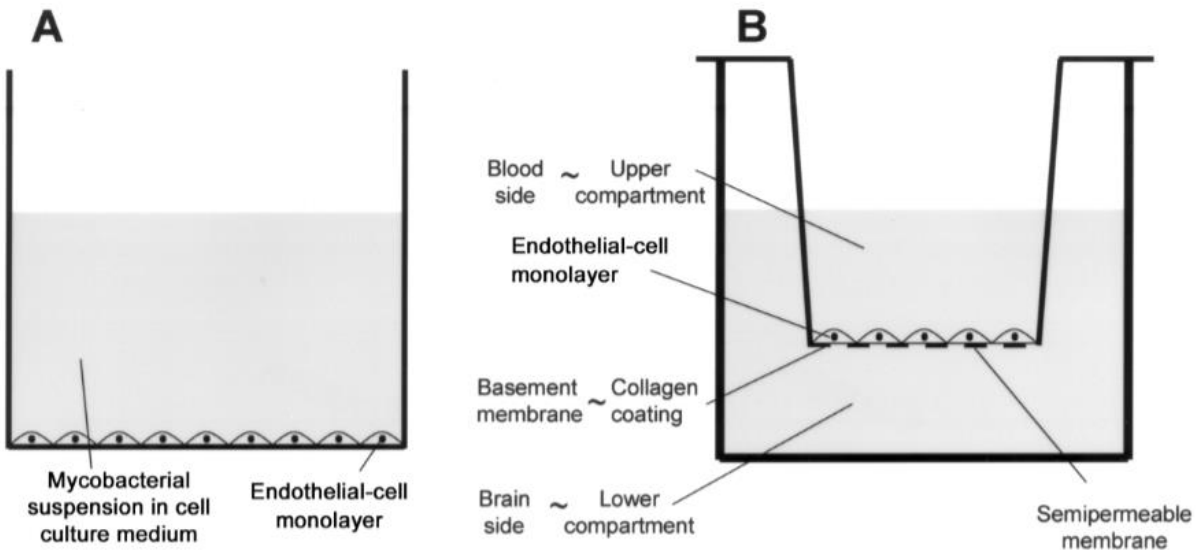

Figure 1.5: (A) Invasion assay: Primary endothelial cells line the bottom of a flask incubated with a mycobacterial suspension of either species. (B) Traversal Assay: infant endothelial cells were grown on trans-well inserts. Separate access was provided for the upper (blood side) and lower (brain side) compartments. The monolayer was incubated with a mycobacterial suspension to induce traversal to the upper compartment. After 48h the contents of the lower compartment was plated to determine the number of mycobacteria that traversed the membrane. Adapted from [40]

1.3.4 Symptoms of TBM

The symptoms associated with TBM vary and are broad, which complicates diagnosis. Many cases of TBM begin with non-descript symptoms and are often diagnosed too late, once brain damage has already occurred [11]. One of the points of differentiation between TBM and diseases with similar etiology is the prolonged persistence of non-specific symptoms [40]. TBM generally develops in the 3 months following primary infection, and pulmonary involvement may play a role in only a small subset of patients [20]. Symptom presentation differs between children and adults with TBM. Adults generally present with non-descript symptoms common to all forms of meningitis, including fever, headaches, neck-stiffness, focal neurological deficits, changes in behaviour and altered consciousness. In these patients pulmonary involvement varies between 30 to 50% [41, 42].

In children, symptoms are similar to those of adults with the addition of abdominal involvement such as nausea and associated vomiting. Neurological symptoms are also slightly altered as headaches are not as frequently observed in children as in adults but seizures are more likely.

However, neurological symptoms depend on disease stage and progression [43, 44]. A particular feature of TBM in children appears to be a close association between TBM and miliary TB. This relationship is thought to exist as a result of extensive haematogenous dissemination in children which increases the likelihood of Rich focus development [20]. Symptoms in both adults and children have been shown to present acutely, with symptoms usually observed later in pulmonary disease often noted much earlier in TBM [40]. Disease progression modifiers such as mycobacterial strain genotype, drug resistance status of the invading strain, HIV and TB co-infection and BCG immunisation status do not consistently influence disease presentation nor disease acuteness.

There are several stages of disease that allow physicians to determine the correct treatment regimen for the patient. This staging system follows Glasgow Coma Scores associated with tell-tale observable traits. Grade I TBM is described as orientated and fully alert without any signs of neurological damage. Grade II is associated with a Glasgow coma score of 11-14/15 with some neurological deficits. Grade III has a Glasgow coma score of <10 and may include focal neurological deficits [45].

1.3.5 Diagnosis of TBM

Central to any TBM diagnostic or treatment regimen is the need for early diagnosis and treatment for better outcomes and patient survival. Owing to the non-definitive symptoms of TBM, diagnosis proves extremely challenging and requires extremely sensitive and specific diagnostic tests. One of the primary indicators for TBM diagnosis is symptom persistence for longer than 5 days [46]. This indicator scale is central to the widely used Thwaites Diagnostic Index which was described as 86% sensitive and 79% specific for diagnosis in Vietnamese adults. Subsequently this indexing system has been used in several other countries with sensitivities ranging between 96-98% and specificities from 68-88% [47–49]. Testing in a variety of different populations exposed the major disadvantage of the Thwaites Diagnostic Index, being the significant number of false positives (12 out of 86) for patients with cryptococcal meningitis thus lowering the sensitivity and specificity to 78% and 43% respectively [50].

Diagnostic methods for TBM vary in effectiveness, time to result and cost, with each technology having specific strengths. TBM is primarily diagnosed by microscopy using Ziehl-Neelsen

staining techniques for acid-fast bacilli in the CSF. However, the results obtained are usually poor, indeterminate and varied where sensitivity rarely exceeds 60% [46, 51]. Nucleic acid amplification tests (NAATs) are commercially available rapid methods of detection. NAATs should only be used in conjunction with other diagnostic methods and cannot be used to rule out TBM [52], due to the low sensitivity of the test at 56% (95% CI: 46%-66%), although specificity is 98%. The major challenge is the acquisition of sufficient volumes of nucleic acid for tests due to the low number of bacilli and the presence of CSF amplification inhibitors [23]. Real-time polymerase chain reaction (rtPCR) technology has led to the development of a GeneXpert assay to simultaneously detect *M.tb* and rifampicin resistance, critical for MDR pTB detection [53, 54]. The Xpert MTB/RIF assay shows great promise with sensitivity values of 80.4%. This assay has yet to be tested on CSF samples and its efficacy in TBM diagnosis is unknown [55]. The TB whole blood test, Interferon-γ release assay (IGRA), can be used to diagnose pTB but cannot differentiate between latent and active TB [56]. These assays are possible diagnostic tools for TBM but low sensitivities (50-70%) have hindered progress for a definitive TBM diagnostic test [55, 57, 58].

Adenosine deaminase (ADA) catalyses the deamination of adenosine to inosine and ammonia [59]. ADA has two isoforms namely ADA_1 and ADA_2, where ADA_2 localises to monocytes and plays a role in immature T-cell activation [60]. ADA_2 is the major component of total ADA concentration observed in TBM and may be a useful diagnostic marker to distinguish between forms of meningitis due to the significant upregulation of ADA_2 levels in TBM compared to other forms of meningitis [61–65]. Cut-off values used affect sensitivities and specificities ranging between 57-96% and 78-87%, respectively [60, 66]. Radiographic diagnosis is often used in conjunction with other tests. However, no radiographic findings can be specifically attributed to TBM to enable definitive diagnosis. The common features of TBM that can be observed by radiological imaging include basal meningeal enhancement, hydrocephalus, and infarctions of the brain parenchyma and brain stem [67]. Diagnosis of TBM using computer tomography (CT) scans through observations of basal meningeal enhancement, tuberculoma or both were 89% sensitive and 100% specific for TBM [68].

1.3.6 Treatment of TBM

Treatment regimens recommended for TBM follow a similar course to that of pulmonary TB, with emphasis placed on early treatment initiation. The course recommended by the Centres for Disease Control and prevention (CDC) and the American Thoracic Society is initiated with a 2 month induction period in which isoniazid, rifampicin, pyrazinamide and ethambutol are used in conjunction [69]. This is followed by a longer period of 7-10 months of isoniazid and rifampicin administration as a maintenance regimen. Due to the rare phenotype of TBM disease, a clinical trial has not yet been conducted to elucidate the correct duration of treatment and optimal drug regimen. Therefore treatment durations, dosages and the drugs used for TBM are all extrapolated from those used to treat pTB [23]. Typical phenotypic presentations before, during and after treatment are shown in Figure 1.6.

The advantage of the drugs in the regimen for pulmonary TB is that both isoniazid and pyrazinamide are readily able to cross the BBB [70]. This is an important consideration, as the BBB has the ability to limit CSF drug concentrations. Rifampicin and ethambutol, in contrast, display reduced CNS penetration despite still playing a role in TBM treatment [71]. The treatment of drug-resistant strains relies mostly on the specific drugs that they remain sensitive to and often leads to an increase in the number of drugs used in the treatment regimen. No guidelines exist for the treatment of drug-resistant TBM and therefore current MDR TBM regimens are extrapolated from the guidelines for pulmonary TB [72].

Several studies have outlined the use of adjunctive corticosteroids in TBM treatment as it is associated with a reduction in mortality, although the mechanism is unknown. Corticosteroids assist in the alleviation of the inflammation associated with TBM and thus may facilitate the transport of drug compounds across the BBB [73, 74].

Recent developments in treatment regimen formulation have led to the inclusion of thalidomide therapy at Tygerberg Hospital in Cape Town, South Africa [75]. Thalidomide, a known TNF-α inhibitor, is included in regimens at a dosage of 3-5mg/kg/day for a duration of 6 months in cases presenting with TB pseudoabscesses and mass lesions. TB pseudoabscesses are thought to form part of immune reconstitution inflammatory syndrome (IRIS) and can be worsened with concurrent HIV infection. Thalidomide stimulates natural killer (NK) cells and leads to the inhibition of lymphocyte apoptosis [76]. Cytokine production of interleukin-4 (IL-4) and IL-5 are

increased whereas interferon-γ (IFN-γ) production is decreased along with CSF TNF-α during thalidomide treatment. Thalidomide use in HIV positive TB patients has been shown to enhance HIV-specific CD8 T-cell activity but has no effect on viral numbers. As thalidomide functions in abscess reduction, its recommended use is only in patients experiencing neurological compromise as a result of TB pseudoabscesses and mass lesions, approximately 80% of participants had a good outcome, with a 3.8% mortality rate, significantly lower than previous mortality statistics [77].

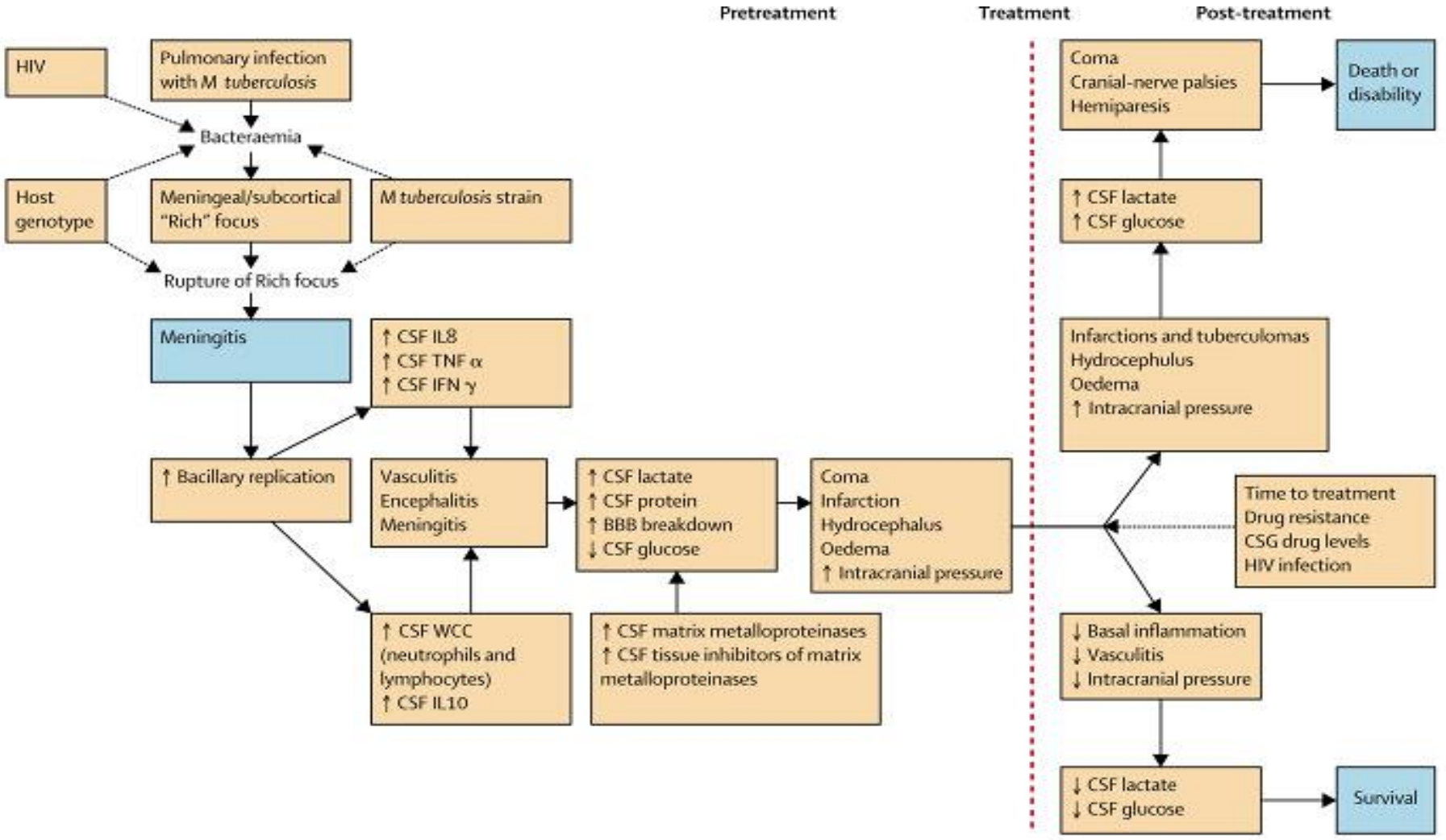

Figure 1.6: Factors influencing TBM presentation and symptoms. Arrows indicate contributions from the environment, host genetics, immune system and bacterial strain. The regulation of variables during treatment and post-treatment phases are also represented showing the outcomes of treatment regimens. Adapted from [78]

1.4 Models used in the study of TBM

Models provide a cost-effective and ethically amenable route of studying disease dynamics and pathophysiology. Each model has advantages and disadvantages to its use and is suitable for different aspects of disease. Disease progression can be studied for the lifetime of the model, their genetic history is known, and knockout models are easy to create [78]. Conclusions can be extrapolated to humans depending on physiological similarities. Several models have been used to study TBM, from mouse models to zebrafish and synthetic models. All have been used to elucidate specific dynamic aspects of the disease.

1.4.1 Rabbit Models

The first animal models used to study TBM were rabbits and guinea pigs [22]. An investigation into the role of TNF-α was performed using intracranial injection of *M. bovis* Ravenal into a rabbit model [79]. An acute inflammatory response was induced in the CSF and allowed for the identification of viable mycobacterial cells. Distinct clinical symptoms of TBM were noted in the rabbit including the identification of granulomas and dissemination to other organs. A sub-acute model was used to evaluate thalidomide treatment response and the efficacy of a recombinant vaccine for the treatment of TBM [80–82]. Investigations with different clinical isolates [83] demonstrated differences in bacillary loads between the strains in the CSF and brain parenchyma and elevated levels of TNF-α and leukocytosis, both of which are observed in human disease. It was concluded that the differences noted between the strains could be attributed to the production of a phenolic glycolipid (PGL), a known virulence factor.

1.4.2 Murine Models

Several mouse models have been used in the study of TBM through the use of intracranial injection to invoke meningeal inflammation and a similar immune response to that seen in humans. BALB/c and DBA/2 mice inoculated with *M. bovis* BCG Montreal to detect microglial activation in the brain demonstrated viable growth of mycobacteria in the meninges and surrounding areas of the brain [84]. Models have also been developed or intracranial inoculation of H37Rv into C57BL/6 mice to elicit an inflammatory response similar to that seen in humans that leads to leukocytosis [85]. However, no neurological defects could be modelled and no granulomas were observed. Intratracheal infection was used to infect BALB/c mice with 3 clinical strains of *M.tb* to observe the disseminatory capabilities of each strain and the related symptoms exhibited by each model

[13]. Inflammation and extensive dissemination in the cerebral parenchyma and surrounding meninges were observed in the TBM clinical strains when compared to pulmonary TB clinical isolates which displayed strain-dependent dissemination.

1.4.3 Zebrafish Models

Zebrafish models have been useful in several diseases due to their genetic similarity to humans (70%) [86]. Dissemination of *M. marinum* in zebrafish has elucidated pathways involved in pathogenesis, with granuloma formation in the meninges occurring in approximately 70% of cases and occasionally in the brain parenchyma as well [10]. Embryos were also used to model infiltration into the CNS through the BBB after inoculation at 3 sites: caudal vein, hindbrain ventricle and brain parenchyma. Each model displayed high bacterial loads in the brain tissue, whether embryos were infected before or after BBB formation. This illustrated that bacteria were able to cross the BBB with high efficiency, although the method remains unconfirmed. Granulomas observed in adult zebrafish were similar to those in humans in that they develop in close relation to the meninges and surrounding brain tissues, providing a useful proxy for modelling early stage TBM infection.

Although each of the above models may reproduce some aspects associated with TBM, no model has been described that cumulatively models disease progression, immune responses and the resultant neurological defects seen in human TBM. Importantly, each of these models is artificially generated through bacterial injection and therefore lack comparability with human disease.

1.5 Immune Response to TBM

1.5.1 Toll-Like Receptors

Human immune responses to TBM are governed by cell-mediated immunity. The process is initiated by recognition of mycobacterial cell wall carbohydrates and proteins via specialised receptors such as the toll-like receptors (TLRs) [87]. TLRs recognise pathogen-associated molecular patterns (PAMPs) and activate signalling pathways to stimulate the innate immune response, cytokine production and the adaptive immune system [88]. Mycobacteria are typically recognised by TLR1, -2, -4 and -6 which require interaction with the adaptor protein myeloid differentiation primary response 88 (MyD88) to facilitate the production of nuclear factor κ B subunit 1 (NFκB). Importantly, lipoarabinomannan and *M.tb* lipoprotein are both recognised by TLR2 [89]. TLR2 is highly expressed on cells of the haematopoietic system such as B and T

lymphocytes and particularly macrophages, monocytes, granulocytes and dendritic cells. Mouse knock-outs have shown that TLR2 is crucial for defence and recognition of *M.tb*, as these mice displayed impaired cytokine and nitric oxide production and failed to mount an effective killing response [90–92]. Immune receptor pathways such as the TLR pathway provide ideal candidates for association studies as polymorphisms in any of the receptor genes or respective adapter protein genes may lead to an ineffective response.

1.5.2 TNF-α

One of the central cytokines in mounting an immune response against *M.tb* is TNF-α. TNF-α induces apoptosis in tumour cells and mediates a large variety of biological activities, such as formation and maintenance of granulomas [93]. The functions of TNF-α display concentration-dependent effects; when present in high concentrations TNF-α has been shown to be toxic to the host causing cachexia (tissue necrosis and wasting) [94]. Conversely when concentrations are low the protective functions are insufficient and host defence is severely compromised [79, 93, 94]. Rabbits inoculated with one of three TB strains induced differential levels of TNF-α. *Mycobacterium bovis* Ravenal induced the highest levels of TNF-α and led to the development of acute meningitis, while rabbits inoculated with strains which induced low levels of TNF-α did not develop meningitis [79]. To confirm that TNF-α was directly related to meningitis development, a nonvirulent strain was transformed with a recombinant murine *TNF-α* gene and the resultant strain induced meningitis. Thalidomide treatment led to the inhibition of TNF-α production and a reduction in leukocytosis and meningeal inflammation in the CSF [79]. Increased levels of TNF-α and IL-1β have been associated with fever, seizures, spasticity and death, possibly due to induction of pro-coagulant production [79]. This in turn causes endarteritis and vessel occlusion, triggering the nervous system sequelae of TBM [95–97].

Leukotriene A4 Hydrolase (LTA4H) regulates the balance between pro- and anti-inflammatory cytokines [93]. A homozygous CC genotype for the *LTA4H* rs17525495 single nucleotide polymorphism (SNP) leads to an anti-inflammatory phenotype which lowers TNF-α to detrimental levels. Conversely TT genotypes lead to a hyper-inflammatory phenotype and high TNF-α concentrations. Heterozygous genotypes maintain the delicate balance between the pro- and anti-inflammatory responses and protects against pTB and TBM in a Vietnamese population [98, 99]. The complex balancing mechanism model is presented in Figure 1.7. Heterozygous advantage is

uncommon and thus suggests that dysregulation of the inflammatory response is linked to TBM susceptibility [100].

Mycobacterial PGL was found to inhibit production of TNF-α and other pro-inflammatory cytokines such as IL-12, IL-6 and Fas-ligand [83]. PGL is a cell-surface lipid that promotes phagocytosis of *M.tb* by host macrophages [101]. It was found that PGL participated in the induction of CNS inflammation through interaction with TNF-α, thereby influencing the virulence of *M.tb* strains [83].

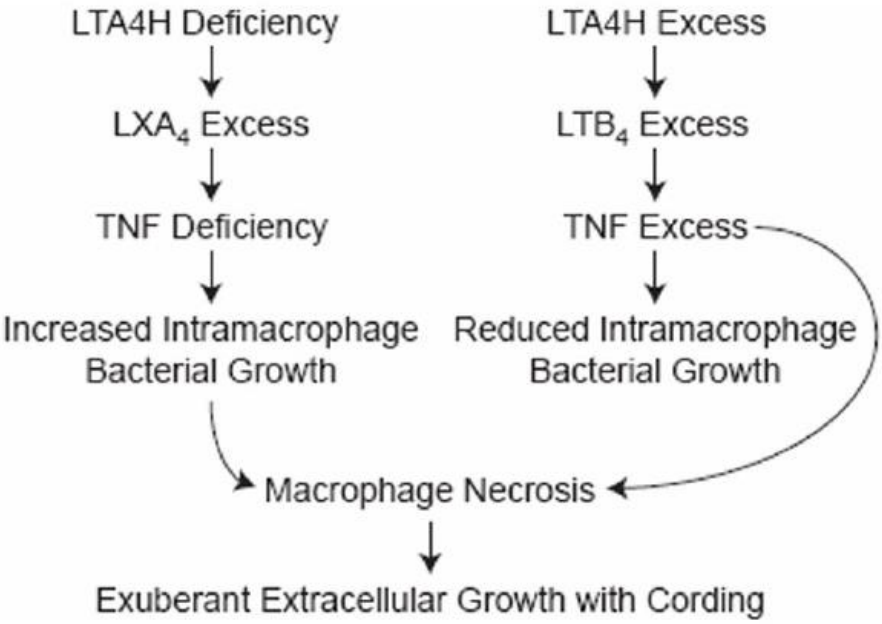

Figure 1.7: A proposed model for the two susceptible genotypes of LTA4H. With the homozygous SNP allele represented on the left side and the homozygous wild-type genotype represented on the right. Adapted from [98]

1.5.3 Immune cells and cytokine production

Cytokine level fluctuations can directly alter the efficacy of the immune response to TBM and affect the course of the disease [18]. Cytokines in the CSF of patients typically include: TNF-α, IFN-γ, IL-1β, IL-6, IL-8 and IL-10, which typically display differential expression in TBM patients and are not typical of other forms of meningitis [102, 103].

Macrophages and microglia are commonly found in the meninges and surrounding blood vessels and have an approximate life-span of 100 days. Macrophages are found ubiquitously in the body, while microglia are found only in the brain [93]. The largest proportion of microglia in the brain

exists in the ramified state whose main function is immune-surveillance [104]. Ramified microglia regain their phagocytic capability by becoming reactive following a pathogen encounter. Both microglia and macrophages produce chemokines such as C-C motif chemokine ligand 2 (CCL2), CCl5, and chemokine C-X-C ligand 9 (CXCL9)/10/11 which recruit T-cells to the site of infection.

Phagocytosis of invading mycobacteria is one of the main functions of both macrophages and microglia [93]. This process is receptor-mediated through cell-surface molecules. The phagosome fuses with a lysosome to release toxic chemicals such as nitric oxide (NO). If the phagocyte kills *M.tb* then antigens will be presented to cluster of differentiation 4 ($CD4^+$) and $CD8^+$ T-lymphocytes increasing the cytokine complement to aid in bacterial clearance.

Pathway analysis has been used to provide insight into the upregulated pathways involved in mounting an immune response against TBM [18]. Upregulated levels of CCL3, CCL4 CCL5, CXCL10 and C-C chemokine receptor 5 (CCR5) proteins were similar to that observed in patients with multiple sclerosis (MS). MS hinges around the demyelination of the CNS, leading to an increase of activated T-cells and macrophages in the CNS inducing leukocytosis and inflammation. This led to a hypothesis that similar treatment regimens may be used in patients with TBM to decrease hyper-inflammation, the aim being to reduce the number of survivors with debilitating neurological sequelae.

1.6 HIV and TB Co-infection

Owing to the rapid spread of HIV, an increase in the incidence of new TB cases has been observed. This comes as a direct result of the immune-compromisation of patients with HIV and has led to poorer prognosis and an increased risk of re-activation of latent TB infections [105, 106]. Healthy HIV-uninfected patients exhibit a lifetime risk of developing TB of between 10% and 20%, which increases in HIV-infected patients to above 20% [107]. HIV-infected patients display an increasingly rapid TB disease progression and likelihood of TBM development of approximately 1 in 3 persons, when presenting with active TB [6, 42, 108]. Co-infection complicates treatment regimens as the risk for relapse in these individuals is greatly increased and subsequently leads to increases in drug-resistance [109]. Co-infection does not affect anti-tuberculosis drug activity despite the immunosuppressed nature of the host immune system [42]. HIV does not alter the presenting symptoms of TBM patients, but instead may alter the number and nature of secondary neurological complications associated with the disease [25]. Co-infected patients display fewer of

the defining pathological features of TBM, with post-mortem examinations finding fewer basal exudates and larger numbers of acid-fast bacteria in the meningeal and cerebral parenchymal regions [42]. Other features include lower leukocyte counts in the CSF due to the targeted attack on $CD4^+$ T-cells by the virus which may explain the greater number of acid-fast bacilli observed in the CSF of co-infected patients, and the increased likelihood of concurrent miliary TB and TBM. HIV infection status has been associated with increased disease severity in patients and thus mortality in these patients may exceed 60% [110, 111].

Studies relating to the complex relationship between HIV and TBM have been few. A study using 96 co-infected and 432 HIV-infected patients was done in Southern Vietnam, showing associations with male sex ($p < 0.001$), age ($p < 0.001$), weight ($p = 0.018$), Glasgow coma score ($p = 0.018$) and peripheral leukocytes and neutrophils ($p < 0.001$ and 0.032 respectively), amongst others [42]. It was also noted that increased concentrations of aspartate transaminase and alanine aminotransferase in co-infected patients had potential as diagnostic markers. Interestingly several groups have found that co-infection with HIV did not affect patient survival nor patients with neurological sequelae [42, 106]. The complex relationship between HIV and TB could be explained by HIV's ability to render TNF-α mediated macrophage apoptosis non-functional [112, 113].

1.7 Host genetics of TBM

Historically, it was thought that TB disease was inherited but following the discovery of the causative bacterium, focus shifted towards the capacity of the host immunophenotype to counteract the pathogen [114]. Subsequent inquests into TB disease aetiology revealed a complex interaction between host, pathogen and environmental factors. Twin and immunological studies have identified a definite host genetic component, with TB heritability ranging from 36% to 80% [3, 4]. Several susceptibility loci for TB have been identified, mainly through candidate gene association studies and genome-wide linkage analyses [115]. Despite the identification of numerous susceptibility genes for TB, such as natural resistance-associated macrophage protein 1 (*NRAMP1*), human leukocyte antigen DR beta 1 chain (*HLA-DRB1*) and *CCL2*, a large proportion of the heritability of this complex disease remains unidentified [14, 115]. In the case of TBM, many genes have been highlighted as susceptibility genes, a large proportion of which function as part of the innate immune system [116].

1.7.1 SNPs associated with TBM

Candidate gene association studies are a popular study design to investigate genetic susceptibility to TBM. Usually genes thought to have a functional effect during *M.tb* infection are investigated. Candidate genes chosen often play central roles in critical immunological pathways associated with infection. There is some evidence for association with TBM for SNPs in the *TLR* and toll-interleukin 1 receptor adaptor protein (*TIRAP*) genes, purinergic receptor (*P2X7*) and cytokine genes, but in all cases were population specific.

1.7.1.1 SNPs in the TLR and TIRAP genes

Polymorphisms in genes encoding proteins that function within the TLR pathway may lead to protein inactivity as they are unable to recognise targets and subsequently induce signal transduction. Previously, *TIRAP* and *TLR2* SNPs have been associated with TB in a Thai but not an African population [117, 118]. The C allele of the *TLR2* g.597T>C SNP, was found to be associated with TB in the Thai population (p = 0.007 [odds ratio (OR) = 2.22, 95% CI: 1.23-3.99]) [118]. However, the association was increased in the TBM group (p = 0.0002 [OR = 3.26, 95% CI: 1.72-6.18]) and was found to be strongest in the presence of miliary TB and higher grade neurological symptoms (p = 0.0002 [OR = 5.70, 95% CI: 1.81-18.0]). The association was found to be highest in individuals with a dominant 597CC homozygous genotype. Polymorphisms in *TLR2* may affect the ability of the receptor to recognise ligands, this could be due to the inability to heterodimerise with either TLR1 or TLR6. Differences in allele frequency between populations may explain the lack of replication of the association in the African cohort [90–92]. An association between the g.597T>C SNP of *TLR2* and the Beijing strain of *M.tb* was strongest of all of the strains tested (p = 0.001 [OR = 1.91, 95% CI: 1.28-2.86]) [116]. The Beijing strain genotype is considered to be one of the most virulent strains due to an intact polyketide synthase 15/1 (*PKS 15/1*) gene and the production of PGL. PGL plays a role in the dissemination process as knock-out $PGL^{-/-}$ mycobacteria were less capable of dissemination than wild-type [83, 119, 120]. The association between the Beijing strain genotype and the *TLR2* SNP demonstrates the first relationship between a bacterial and host genotype in TB.

The *TIRAP* gene encodes a critical adaptor protein in the TLR pathway. Its biological function of signal mediation from TLR1, -2, -4 and -6 makes it a plausible candidate gene. Several SNPs in the *TIRAP* gene have been investigated. The T- allele of the g.539C>T SNP (S180L) is protective

against pneumococcal disease, malaria, and pulmonary TB in Caucasian, South American and African populations [121, 122]. *TIRAP* SNP g.558C>T was associated with TBM patients ($p < 0.001$ [OR = 3.02, 95% CI: 1.79-5.09]) [88]. Analysis of IL-6 production *ex vivo* showed an association between individuals with a 558TT genotype and a decrease in IL-6 production, 72.7 pg/ml [95% CI: 6.7-91.2] compared to the 558CC genotype, 329.9 pg/ml [95% CI: 155.3-965.3] when stimulated with PAM2Cys-SKKK (PAM2). These results show that TB cases carrying the 558TT genotype display inefficient IL-6 production, contributing to uncontrolled bacterial growth and eventual dissemination.

Screening of the open reading frame (ORF) of the *TIRAP* gene in two South African populations, a Xhosa and a mixed ancestry population, found similar results to previous studies [88]. However, the association of g.558C>T with TBM was restricted to the mixed ancestry population in both heterozygous ($p < 0.05$), and homozygous 558TT ($p = 0.02$) genotypes [117]. No significant differences in cytokine production were observed during Lipopolysaccharide (LPS) stimulation of the TLR4 pathway. Instead, TNF-α levels differed between g.548G>C, a novel SNP identified, and g.589G>A which showed increased TNF-α than wild-type individuals in the Xhosa group, $p < 0.05$. Differences between the results of the two studies could be attributed to differences in allele frequencies between population groups. Differences in cytokine production can be attributed to different ligands used for TLR stimulation, as Hawn et al. (2006) used PAM for stimulation of the TLR2 pathway, whereas Caws et al. (2008) used LPS, a known TLR4 ligand [94, [117]. TLR2 is the main recognition receptor for *M.tb*, which suggests that the effect of the g.558C>T SNP is specific to TLR2 signalling and may therefore influence the occurrence of TBM in affected patients.

1.7.1.2 SNPs in the $P2RX_7$ gene and their effect on ATP-mediated killing of mycobacteria

The $P2X_7$ receptor ($P2RX_7$) is expressed on the surface of human macrophages and plays a role in mycobacterial defence [123]. $P2RX_7$ is activated by extracellular ATP to induce its killing activity upon phagocytosed intracellular mycobacteria, acting independently from reactive oxygen and nitrogen-mediated killing systems [124, 125]. Activation of $P2RX_7$ opens cation-selective channels mediating calcium (Ca^{2+}) influx, this activates a caspase cascade resulting in mycobacterial apoptosis and the activation of phospholipase D (PLD) [126–128]. The role of PLD is to induce fusion between the phagosome and the lysosome containing cytotoxic enzymes [125,

129]. Several SNPs have been found in the *P2RX7* gene, with the most common of these being g.1513A>C [130]. The consequential amino acid substitution affects many of $P2RX_7$ functions, including the ability to create cation influxes. G.1729T>A displays a functional hindrance by preventing the receptor trafficking to the cell surface [131]. g.946G>A prevents ATP binding required for the activation of downstream caspase cascades [132]. The g.151G>T SNP results in a null allele [133]. The cumulative effect of these polymorphisms affects the responsiveness and activity of $P2RX_7$ [134, 135].

g.1513A>C was found to be associated with all forms of extrapulmonary TB (including TBM patients), but not pulmonary TB ($p < 0.01$ [OR = 3.8, 95% CI: 1.6-9.0]) [123]. Allelic analysis found that individuals who were homozygous for the 1513C allele displayed abrogated ATP-mediated killing upon phagocytosed mycobacteria whereas heterozygotes displayed severely limited killing capacity. The authors posited that loss of $P2RX_7$ could lead to unhindered spread of mycobacteria and subsequently higher bacterial load. A meta-analysis of g.1513A>C and the NG_011471.2:g.-762T>C promoter SNPs, previously found to confer susceptibility in a Gambian population [136], showed that A1513C was associated with TB ($p < 0.00001$ [OR = 1.44, 95% CI: 1.23-1.68]) in 1044 cases and 1286 controls across 6 studies. NG_011471.2:g.-762T>C was not associated with TB ($p = 0.97$ [OR = 1.01, 95% CI: 0.70-1.44]) in 857 cases and 1068 controls. Similar results were observed between g.1513A>C and TB [137]. Sub-group analysis showed stronger association with extrapulmonary forms ($p < 0.001$ [OR = 1.39, 95% CI: 1.16-1.66]) with the 1513CC genotype found to completely ablate the ATP-mediated killing response ($p = 0.012$ [OR 1.58, 95% CI: 1.13-2.22]). Contrary to this, no association was found between the g.1513A>C SNP and either pulmonary or extrapulmonary TB forms ($p = 0.275$). This result highlights the effect of allele frequency differences between different population groups upon association results [138].

1.7.1.3 SNPs in IFN-γ

Interferons are proteins that are produced and exported from immune cells in response to pathogen recognition. IFN-γ is produced by T-helper 1 cell (Th1) lymphocytes and plays a role in macrophage activation [139]. Several polymorphisms in the *IFN-γ* gene have been studied, such as the g.874A>T (rs2430561) SNP, the T-allele of which was shown to confer a protective advantage against pulmonary TB in a South African population ($p = 0.0055$ [OR=1.64, 95% CI:

1.16-2.30]) [140]. Meta-analysis concluded that the 874T allele confers the protective advantage observed [140, 141]. A study in a Chinese Han paediatric female population group investigated 3 SNPs of the *IFN-γ* gene, namely g.1616C>T (rs2069705), g.874A>T (rs2430561) and g.3234C>T (rs2069718) for association to extrapulmonary TB [142]. Their results indicated that the 874A and 3234C alleles were both associated with extrapulmonary TB (p = 0.039 [OR = 1.53, 95% CI: 0.92-2.54]). Additionally, the two SNPs were found to be in linkage disequilibrium (LD) (r^2 = 0.987). Haplotype analysis showed the TTT haplotype was less prevalent in cases compared to controls, indicating a protective advantage (p = 0.005 [OR = 0.46, 95% CI: 0.26-0.79]). Individuals carrying the TTT haplotype are protected against TBM disease due to an intact IFN-γ response and consequent macrophage activation.

1.7.1.4 SNPs in IL-4

IL-4 is an anti-inflammatory cytokine that is produced by Th2 cells during immune responses against *M.tb*. IL-4 has been shown to exert its anti-inflammatory effect through the down-regulation of IFN-γ and other Th1 related responses, which leads to deleterious effects in the host if unregulated [143, 144]. Two SNPs in *IL-4* were found to exert protective effects against TBM and other forms of extrapulmonary TB, rs2243268 (p=0.005, OR=0.59 [95% CI: 0.41-0.85]) and rs2243274 (p=0.004 [OR=0.58, 95% CI: 0.40-0.84] [143]. The AA and GG protective genotypes respectively led to significantly lower production of IL-10 upon stimulation with inactivated H37Rv (p=0.045) and thus had a weakened anti-inflammatory response (IL-4 levels were unmeasurable). Efficient regulation of an anti-inflammatory response is central to mounting a sufficient response to TB infection. The proper maintenance of cytokine homeostasis may be key to controlling the dissemination observed in early TBM development.

1.7.2 Strain Genotype Effects

Differences in disease severity and patient outcome have been observed between *M.tb* strains, leading to investigations into the link between the most pathogenic strains and increased disease severity. Beijing TB strain isolates are considered to be among the most virulent, thus it is unsurprising that they have been associated with increased disease severity and mortality rates in HIV-TBM co-infected patients. Several studies have investigated the associations between strain genotypes and severity or likelihood of disease development in extrapulmonary TB [16, 42, 53, 109, 116, 145–147]. Authors have hypothesised that some *M.tb* strains could possess increased

disseminatory capacity because of the interplay between the strain and host genotype related to polymorphisms in host *TLR* genes [116]. They compared the host and mycobacterial genotype between 187 TBM patients, 237 pTB patients and 392 controls from a Vietnamese population. Interestingly a protective association was found between the Euro-American lineage of *M.tb* and TBM ($p = 0.009$ [OR = 0.395, 95% CI: 0.193-0.806]). This suggests that the Euro-American strains are less capable of causing disseminated disease than other more virulent strains. In accordance with this, patients infected with the Euro/American strain displayed comparatively shorter disease courses ($p = 0.039$ [OR = 0.973, 95% CI: 0.948-0.999]). In further support of this theory, an association was found between the C allele of the g.597T>C *TLR2* SNP and strains of the Beijing genotype, therefore patients with 597CC genotype were more likely to have TBM caused by a Beijing genotype strain ($p = 0.004$ [OR = 1.57, 95% CI: 1.15-2.15]) [116].

In support of the results found by Caws et al. (2008), an independent study by Kong et al. (2007) studying a population in Arkansas, USA found that patients infected with Beijing strain isolates were approximately 3 times more likely to have extrapulmonary disease as other patients infected with non-Beijing isolates (OR = 2.85 [95% CI: 1.33-6.12]) (p-value unavailable) [116, 146]. An association was also found between the Beijing strain genotype and disease in patients younger than 15 years old ($p = 0.018$ [OR = 4.47, 95% CI: 1.10-18.17]) [147]. Contrary to these results, two studies from South Africa found no association between invading strain genotypes and clinical features of TBM disease in their respective study populations [16, 148]. This could be due to population or design differences between the studies.

Investigations into the relationship between HIV-TBM co-infected patients and specific strains aim to elucidate whether one strain in particular is more associated with co-infected TBM disease than any other [42, 53, 109]. HIV-TBM co-infected patients in a Vietnamese cohort have higher incidences of MDR-TB ($p = 0.034$ [OR = 12.35, 95% CI: 1.23-125.00 and mortality ($p = 0.001$, OR = 4.24 [95% CI: 2.10-8.54]), compared to HIV un-infected TBM patients [72, 109, 149]. This may be due to host immunosuppression as a direct result of co-infection with HIV and subsequent treatment regimen complication. The Beijing strain was associated with both HIV-positive status ($p = 0.0.16$ [OR = 2.95, 95% CI: 1.38-6.44]) and acquisition of drug resistance ($p < 0.001$ [OR = 3.34, 95% CI: 1.87-5.95]) [109]. This suggests that Beijing isolates may more readily cause TBM disease in HIV-infected TBM patients and have a higher propensity for acquiring drug resistance

mutations independent of host HIV status. HIV further complicates TBM infection due to the reduction in $CD4^+$ T-cells, diminishing the innate immune system, increasing patient susceptibility to *M.tb* infection and dissemination likelihood due to high bacterial loads [150, 151]. In contrast, other studies did not find that HIV-TBM co-infection altered treatment response of patients despite marked immunosuppression [42]. A study by Tho et al. (2012), examined the combined effects of drug resistance and strain genotype on HIV-TBM co-infection patient survival [53]. Isoniazid resistance showed the strongest association with mortality of the patient ($p = 0.005$ [OR = 1.78, 95% CI: 1.18-2.66]) and multidrug resistance was strongly associated with patient mortality in all 8 patients infected with MDR TB ($p < 0.0001$ [OR = 5.21, 95% CI: 2.38-11.42]). These results show that drug resistance status of the strain, host and strain genotypes and patient HIV status are all important determinants of patient survival and thus must be considered during treatment administration.

Scope of the Thesis

The research presented in this thesis investigates the role of genetic variants of differing frequency in the susceptibility to TBM in the South African Coloured (SAC) population. This was done using both exome sequencing gene-set based association analyses and SNP genotyping array technologies to conduct a genome-wide association study (GWAS). In light of this, we hypothesise that multiple common variants of moderate effect size are more likely to influence TBM susceptibility than rare variants of large effect.

The SAC are a 5-way admixed population with ancestral contributions from African, European and Asian population groups. An overview of the population structure, the prevalence of TBM in the population and the number of participants used per study subsection is presented (**Chapter 2**).

The aim of whole exome sequencing (WES) experiments is to study only SNPs in protein coding regions of the genome, because these may alter protein function. An exome sequencing study of 10 cases and 10 controls is described with the resultant sequence data analysed using Sequence Kernel Association Test (SKAT) gene-set association tests. Two SKAT tests are described, firstly using only SNPs of rare to low frequency and secondly assessing the cumulative effect of both common and rare SNPs on TBM susceptibility in the SAC (**Chapter 3**).

Assessment of common SNPs in the SAC population was done using a second technology, that of the Illumina® Multi-Ethnic Genotyping Array (MEGA) for SNP genotyping. SNP selection included functionally relevant SNPs and those previously associated with asthma and other respiratory diseases such as TB. Amongst these, select genome-wide tag SNPs were selected to increase the power for detection of association using exome sequencing content from over 12 000 individuals. Our first GWAS included 114 TBM patients and 395 healthy controls to assess the role of common SNPs in TBM susceptibility. A second GWAS was done with 114 TBM cases and 382 pTB cases to determine SNPs that increase susceptibility to dissemination of *M. tb* to the CNS for TBM development (**Chapter 4**).

Concluding remarks on the results of the study as a whole are presented in **Chapter 5**. Limitations of the study design are discussed and their possible effects discussed. Candidate genes found during exome sequencing and the GWAS highlight the need for further investigation.

Study Participants Chapter

Chapter 2: Contents

Chapter 2: Study Population and Participants

2.1 Study Participants

The Western Cape Province has one of the highest TB incidence rates in the world, approximately 1033 per 100 000, as measured in 2011 [152]. This province also has one of the highest TBM incidence rates in the world, 31.5 per 100 000 in children younger than one year, as measured in 2006 [11]. The major ethnic group in the Western Cape is the admixed South African Coloured (SAC) population (48.8%) [153]. Population genetics analyses have revealed a complex five-way admixture for the SAC which includes contributions from European, East Asian, South Asian, Bantu-speaking African and KhoeSan populations [154]. To account for the possible effects of multiple contributing ancestral populations, it is necessary to correct for population stratification in our genetic analyses [155].

2.1.1 Ethics approval

This study was approved by the Health Research Ethics committee of Stellenbosch University (project numbers 95/072 and N09/07/185). Written informed consent was obtained from all adult study participants and legal guardians. Assent was obtained from minor children.

2.1.2 TBM study participants

HIV-negative TBM patient samples were collected and stored as part of an ongoing recruitment in collaboration with Paediatrics and Child Health, Tygerberg Children's Hospital, Cape Town, South Africa since 1991. Approximately 3 TBM cases are diagnosed per month. Diagnosis is classified into two sub-categories "Proven TBM" and "Probable TBM". Proven TBM cases were diagnosed when *M.tb* is able to be isolated from the CSF. Probable TBM cases are cases that exhibit clinical evidence of meningitis with concurrent hallmark CSF findings or imaging evidence, in addition most likely alternative diagnoses were able to be ruled out. Population and language information was collected for each individual in the study. The TBM samples used in this study included 96 SAC and 30 Xhosa individuals. Sample characteristics concerning sex, age and ancestry proportions are shown in Table 2.1.

2.1.3 Pulmonary TB patients

Pulmonary TB patients used in the study self-identified as part of the SAC population, and were recruited from two metropolitan areas of Cape Town. These areas were selected due to the high

TB incidence (1340 per 100 000) and low HIV prevalence (approximately 2% of the population) at the time of sampling [156, 157]. All study participants were HIV negative and unrelated. Diagnosis of pTB was determined through bacteriological confirmation using either smear and/or culture methods for positivity.

2.1.4 Healthy control individuals

Control samples were collected from the same metropolitan areas as the pTB cases and therefore share the same environmental and socio-economic circumstances. Controls were over the age of 18, HIV-negative and unrelated to one another or to the cases. Additionally, control samples were defined as individuals who had never had a case of active TB in their lifetime, however, this does not exclude the possibility of latent infection. Positive tuberculin skin tests (TST) were observed in 80% of children older than 15 years in this population, indicating significant exposure of control individuals to *M.tb* [158].

Table 2.1: Sample characteristics of the study including sex, age and ancestry proportions per study arm.

		Study		
		Exome Sequencing	GWAS TBM vs. Controls	GWAS TBM vs. pTB
Cases	Number of Cases	10	114	114
	Nr Males (prop)	5 (0.50)	59 (0.52)	59 (0.52)
	Age (mean ± SD)	5.1 ± 3.93	5.24 ± 4.86	5.24 ± 4.86
	African San [IQR]	0.27 [0.24-0.35]	0.29 [0.22-0.38]	0.29 [0.22-0.38]
	African non-San [IQR]	0.42 [0.31-0.74]	0.36 [0.24-0.65]	0.36 [0.24-0.65]
	European [IQR]	0.09 [$1x10^{-5}$-0.14]	0.04 [$1x10^{-5}$-0.10]	0.04 [$1x10^{-5}$-0.10]
	South Asian [IQR]	0.08 [$1x10^{-5}$-0.14]	0.19 [0.05-0.26]	0.19 [0.05-0.26]
Controls	Number of Controls	10	395	382
	Nr Males (prop)	5 (0.50)	118 (0.30)	212 (0.55)
	Age (mean ± SD)	49.65 ± 12.79	30.88 ± 13.10	36.32 ± 11.04
	African San [IQR]	0.31 [0.27-0.45]	0.25 [0.19-0.34]	0.31 [0.20-0.40]
	African non-San [IQR]	0.20 [0.15-0.24]	0.27 [0.19-0.38]	0.24 [0.15-0.36]
	European [IQR]	0.24 [0.22-0.29]	0.12 [0.05-0.20]	0.17 [0.12-0.22]
	South Asian [IQR]	0.11 [0.08-0.15]	0.25 [0.18-0.31]	0.15 [0.10-0.21]

IQR = Interquartile range
SD = Standard deviation
Prop = Proportion
GWAS = Genome-Wide Association Study
To avoid linear dependency in the data, the East Asian component was not added to the model
P values reflect the significance of the association of each factor with TB, adjusted for the other factors

2.1.5 DNA extraction from blood

DNA was extracted from blood according to the Nucleon BACC Genomic Extraction Kit protocol (Illustra, Buckinghamshire, UK). Concentration and purity assessments of DNA were determined using the Nanodrop 2000c and Nanodrop 2000/2000c software (ThermoFisher Scientific Waltham, MA, USA).

Exome Sequencing Chapter

3

Chapter 3: Contents

Chapter 3: Exome Sequencing

3.1 Introduction

The human genome has approximately 3 billion nucleotides of which protein coding exonic regions constitute an estimated 1.22% [159]. The aim of WES experiments is to sequence only protein coding regions of the genome. The hypothesis behind this study design is that SNPs in exonic regions directly affect produced proteins and thereby its function. The approximately 20 000 genes in the human genome are made up of 233 785 exons which are targeted during library preparation [160]. The exome therefore represents a subset of the genome in which SNPs are more likely to exert a direct impact on protein function, especially where non-synonymous amino acid changes are induced.

Preparation for WES involves the construction of a sequencing library with fragmented genomic deoxyribonucleic acid (DNA) as input. DNA shearing is random but size-specific (~150 base-pairs (bp)) to limit off-target reads as most exons are less than 200bp long (Figure 7) [160]. Sequence-specific adaptors are ligated to each fragment and is followed by library enrichment and capture of exonic portions of the genome [161]. This process requires the hybridisation of designed sequence-specific exonic probes to the fragmented DNA. Hybridised complexes of exonic portions are then isolated through biotin-streptavidin based capture isolation in the presence of a magnetic field. The target DNA is then eluted and amplified post-hybrid-selection and sequenced (Figure 3.1).

Despite exome sequencing being a sensitive tool, its application to complex disorders benefits from targeted study design strategies. Such designs involve sampling of rare or extreme phenotypes of disease for sequencing. This approach has been validated through the discovery of novel candidate genes for paediatric inflammatory bowel disease and chronic *Pseudomonas aeruginosa* infection in patients with cystic fibrosis [162–164]. Improvements in WES technologies to improve data accuracy and library preparation are ongoing, but a large hurdle in terms of library probe design remains the identification of protein-coding exonic regions in the genome. Initial probe designs centred on high confidence regions but as the capacity and accuracy of the technology increased so has the number of targeted regions included in contemporary library preparation kits [165].

WES has been previously demonstrated to be a powerful tool in rare disease research through the discovery of a novel gene implicated in a rare Mendelian disease, Miller syndrome [159]. WES allows for the identification and characterisation of rare SNPs in the comparatively low sample sizes found in rare disease research, and provides a smaller, focussed dataset for analysis. WES provides a viable research tool for investigation of rare to low frequency SNPs which are thought to be of large effect size and cumulatively contribute to the overall disease phenotype and rare disease severity [166].

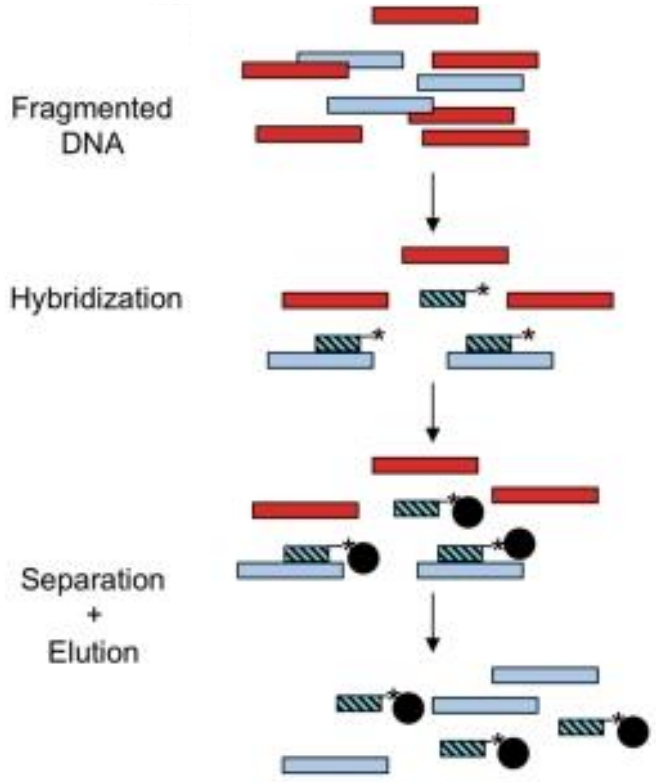

Figure 3.1: Overview of exonic sequence capture through hybridisation with biotinylated streptavidin beads. This is done during library preparation for exome sequencing. Red bars = Intronic genome portions, Blue bars = Exonic regions which are captured by probes (striped bars). Adapted from [161]

3.1.1 Sequence Kernel Association Test (SKAT)

GWAS investigate common variants (frequencies > 1% in the study population) that often display relatively small effect sizes with regards to disease susceptibility [166]. Findings from GWAS have not been able to explain the heritability of TB disease. Additionally, GWAS to date have

displayed a very low level of replication across population groups. This lack of reproducibility may be attributed to differences in allele frequencies between population groups [167]. Research focus has shifted towards investigations of rare SNPs of predicted large effect size, with frequencies in the range of 1-5% (low) and below 1% (rare) [168–172]. Rare variants are relatively uncharacterised in the context of GWAS, as many are removed from analyses by minor allele frequency quality control (QC) measures. GWASs also require large sample sizes for adequate power to detect associations [173]. Rare SNPs may play a role in TBM susceptibility, especially given the incidence of TBM disease in the population, which may point to SNPs of low frequency being causal.

The cumulative effect of rare SNPs in a single gene or across functionally similar genes in a pathway may impact upon the susceptibility of individuals to disease. Given recent population expansions it is plausible that large numbers of functionally relevant rare SNPs account for a proportion of phenotypic variation in the population [174]. Evidence is provided by disorders that are predominately monogenic, where the additive effects of rare SNPs in the same gene were associated with the disease [175, 176].

Rare SNP association analyses face a major hurdle during study design, namely the large sample sizes required to detect the subtle variations in frequency between case-control groupings. To circumvent this issue when testing SNPs individually, studies have ‘collapsed’ SNPs into genetic regions and testing these regions for frequency differences [177]. Using this method, the multiple testing burden is reduced for these tests as independent tests are not performed on a per-SNP basis. Various algorithms have been developed to accomplish a gene-set based analysis such as Sequence Kernel Association Test (SKAT), Cohort Allelic Sums Test (CAST) and the C-alpha test [177–179]. CAST is defined as a burden test, a cumulative association test for rare SNPs that tests the association of overall SNP burden in gene regions with disease susceptibility. Burden tests collapse genotypic information for groupings of rare SNPs into singular genetic scores. The genetic score calculated is then tested for association with the trait of interest. These tests are powerful when the majority of SNPs are causal and exert effects in the same direction. The disadvantage is their lack of power when effects are balanced in directionality or the proportion of causal SNPs is low [171]. SKAT and the C-alpha test on the other hand are termed variance-component tests, which test the variance of genetic effects between regions. This increases the power in studies where SNP

directionality is balanced and a small proportion of SNPs are likely to be causal, but power is decreased when the majority of the SNPs exert effects in the same direction [171].

In order to optimise the power achievable by studies investigating rare SNPs, SKAT-O was developed: an optimised version of SKAT [180]. SKAT-O is more tolerant of the proportion of SNPs that are causal in datasets and also to bi-directionality of SNPs. This is achieved through the combination of both the variance-component testing of SKAT and burden tests to create an optimised test which is tested across a series of values for ρ to select the ideal test per gene-set. The optimised nature of the tests makes them an attractive option as little is known of the underlying genetic architecture prior to testing.

Sequencing-based studies represent a unique opportunity to assess the relative contribution of both common and rare SNPs simultaneously. Previous gene-set based association tests, including SKAT tests, have been able to assess the cumulative effects of SNPs of varying frequency. The focus of these tests was biased to up-weight the effects of rare SNPs and consequently down-weight those of common SNPs [181]. In doing so, considerable losses in power were noted when common and rare SNPs in the same region cumulatively influenced susceptibility to certain traits. Considering that the contribution of rare and common SNPs to disease risk is not known prior to testing for any possible trait locus, it is not surprising that power losses were observed. It is likely that SNPs of varying frequency cumulatively contribute towards gene loci conferring susceptibility to a particular trait. For this reason, omnibus testing procedures such as SKAT Common Rare enable the simultaneous assessment of both common and rare SNPs to determine their contribution to overall disease susceptibility at loci of interest. SKAT Common Rare testing involves the partitioning of SNPs into frequency-based bins and then testing bins separately to determine their relative association in loci of interest. These scores are then combined across genetic loci using a multivariate Hotelling's T-Square statistic. Through this combination method, an overview of the cumulative effects of SNPs of varying frequency in genetic loci can be obtained. This testing procedure is therefore attractive for exploration into disease pathogenesis studies with the aim of identifying candidate genes for future study.

3.1.2 Aims

1. To exome sequence 20 individuals, 10 TBM cases and 10 healthy controls, from the SAC population to discover novel SNPs.
2. To determine if rare SNPs grouped into gene-sets are associated with TBM using the SKAT-O test and thereby identify candidate genes for future study.
3. To assess the cumulative effect of both common and rare SNPs on TBM susceptibility using the SKAT Common Rare test and thereby identify candidate genes for future study.

3.2 Materials and Methods

3.2.1 Sample selection for exome sequencing

Sample selection for exome sequencing involved the selection of 20 unrelated individuals, 10 TBM cases and 10 healthy controls. Cases and controls were sex matched (Table 2.1). Criteria for selection included DNA concentration (>200ng/µl) and quality metrics (260/230 ratio ≤ 2 and 260/280 ≥ 1.7) as determined by Nanodrop, a nucleic acid quantitation method requiring micro-volumes of sample. Age selection criteria introduced a bias as patient samples were much younger than controls due to the age of onset of the disease. Control samples were selected for older age as the risk of TBM development decreases as age increases. Samples were diluted to 200ng/µl in 30µl Tris-EDTA and stored in Eppendorf® Safe Lock Tubes (Eppendorf AG, Hamburg, Germany) and stored at 4°C during transport. Samples were randomised by sex and disease affection status during sequencing to avoid batch effects.

3.2.2 Library preparation parameters

The input genomic DNA was sonicated to randomly fragment the DNA to a size of 150bp [182]. Sonicated DNA quality and size assessment was done using the Agilent Bioanalyser 2100 and DNA 1000 chip and reagent kit (Agilent Technologies Santa Clara, CA, USA). Targeted enrichment for exonic sites was done using the Nextera XT® enrichment kit (Illumina San Diego, CA, USA) which targeted >20 000 genes between 40 to 60x read depth. Library preparation for the Illumina Nextera XT® library (Illumina San Diego, CA, USA) used 6µg of DNA and was performed according to manufacturer's specifications. Paired-end sequencing was performed on the Illumina® HiSeq 2500 (San Diego, CA, USA) at Christian-Albrechts University of Kiel (CAU sequencing Kiel, Germany).

3.2.3 Data pre-processing and QC

Sequencing data was received from the CAU sequencing facility in .fastq format. Reads were mapped to hg19 of the human genome downloaded from the UCSC genome browser [183]. The FASTA formatted hg19 reference was indexed prior to mapping using the Burrows-Wheeler Aligner (BWA) [184]. Indexing of the reference genome allowed for alignment of forward and

reverse .fastq reads to the hg19 reference using BWA-mem. Aligned forward and reverse reads were then combined into a .sam file, sorted by chromosomal mapping coordinates. The .sam file was compressed into a .bam file which was indexed during the process.

In order to obtain a graphical overview of the data, Integrative Genomics Viewer (IGV) was used to view the sorted and indexed .bam file [185]. Using IGV, read mapping quality was assessed along with other metrics such as coverage and read depth. Following graphical read quality assessment, duplicates were marked using picard-tools (Broad Institute, http://picard.sourceforge.net/) and removed to mitigate the effects of polymerase chain reaction (PCR) amplification bias [186]. Mapping statistics such as quality score distributions were assessed using bamtools and picard-tools [187]. To assess a panel of QC metrics including sequence quality, GC content, adapter content and overrepresented sequences, FastQC [188] was run on the .bam file thereby removing adapter sequences.

3.2.4 SNP filtration and prioritisation

Next generation sequencing (NGS) technologies involve the synthesis of short reads of approximately 150 base pairs (bp) with repeated sequencing of regions of interest, resulting in multiple sequencing reads of each position [189]. NGS technologies display a higher error rate than Sanger sequencing, necessitating multiple reads for each position in the exome to ensure data confidence. The exome sequencing portion of the study aimed to assess the role of low frequency SNPs in association with TBM susceptibility. Rare SNPs are defined as those below a minor allele frequency (MAF) cut-off of 0.01 in the population [190]. Filtration methods were aimed at rare variant prioritisation. A single .vcf (variant call format) file contained all called SNPs for all samples. Annotation for frequency information and SNP-specific gene regions amongst other functional annotations were added using a web version of Annovar (wAnnovar) [191]. SNP information for 107 554 SNPs across the 20 sampled individuals were annotated. An overview of SNP prioritisation procedures is summarised in Figure 3.2

The annotated vcf file was filtered to prioritise rare SNPs with a MAF of 0.01 and below. This filtration was primarily based upon the annotated MAF provided by the 1000 genomes project [192]. Second tier MAF filtration was performed using annotations from the Exome Sequencing Project (ESP6500si) which consists of over 6 000 healthy individuals and study participants with

several diseases from different populations [193]. Final MAF filtration was done using frequency information from the Exome Aggregation Consortium (ExAC) [194]. Based on the criteria mentioned above, MAF filtration removed 17 836 SNPs from the dataset. Further filtration for retention of non-synonymous variants was done based on the premise that SNPs which change amino acid sequence will be more likely to affect protein function [195]. Frameshift SNPs were retained as these affect codon usage and may disrupt protein structure and function. SNPs that induced stop codon gains or losses, splice-site SNPs and insertions and deletions were retained during filtration, during which 53 419 SNPs were removed (Figure 3.2).

Conserved sequence sites provide a useful resource for the investigation of SNPs, as evolutionarily conserved sites are most likely to be functionally relevant [195]. Less conserved sites in the genome tend to be more polymorphic and SNPs in these regions are more likely to be tolerated in terms of retaining protein structure and function compared to those in evolutionarily conserved regions. As conserved sites are less polymorphic, variants in these regions are likely to be of low frequency and therefore rare in the population. The combination of SNP frequency and the highly conserved nature of the SNP's location point to increased effect size [196]. Therefore, filtration for the retention of conserved sites was performed based on PhyloP [197] and GERP++ scores [198], both measures of conservation . SNPs with negative scores (n = 19 108 SNPs) from both conservation annotators were removed.

The chromosome numbers and related base-pair positions specific for each SNP in the filtered and annotated vcf were used as a reference for extraction of SNPs of interest from the original vcf file using vcftools 0.1.12b [199]. The output was converted to PLINK format using vcftools. PLINK v1.07 was used to convert the resultant map/ped format files into binary bed/bim/fam formats [200].

3.2.5 SKAT-O Analysis

SKAT-O analysis was performed to assess the rare-variant burden across all genes containing rare SNPs identified using exome sequencing. The Annovar and related PLINK files produced during SNP prioritisation were used during creation of the gene-set file. Gene and SNP information were extracted by scripts written using the R programming environment (available from www.r-

project.org) to create a file of SNP IDs grouped into sets according to their corresponding genes [201]. Using the R script, duplicate entries and those with missing information were removed.

Using the genotype-containing .bim file produced during SNP prioritisation in conjunction with the gene-set file, the SKAT-specific SSD and SSD.Info files were created through the use of incorporated R functions in the SKAT R package [179, 202]. Genotypes from the PLINK-formatted .bim file were stored in a binary SSD file. Information concerning the number of gene sets, the number of SNPs per gene and sample structure of the analysis were stored in the SSD.Info file. As a result of filtration and prioritisation procedures, 8 322 genes (SNP sets) and 16 728 SNPs were included in the analysis.

Prior to running SKAT-O, a null model was created to aid in the determination of p-values in the analysis. The null model was constructed to incorporate dichotomous phenotypes of TBM patients and healthy controls regressed against the study covariates. A single covariate of gender was incorporated into the model. During the construction of the null model, only sex was included as a covariate, as including any of the others in conjunction with sex resulted in complete separation of the data. Complete separation is defined as when the outcome variable separates a single predictor variable or a combination of predictor variables completely and correctly allocates all observations to their respective groups [203]. Complete separation was likely due to the small sample size (n = 20) used in the study and therefore insufficient inter-sample variability was observed in the data to allow for the inclusion of further covariates. This could introduce some bias as the data is not fully corrected for confounding factors such as ancestral populations or age. Age represents a significant stratifying factor in this study, as the patient samples used were very young, given the age of onset of TBM, and these were compared to controls who were the oldest of all available control samples. This was done as control individuals of older age were less likely to develop TBM [204]. To improve the accuracy of p-values estimated under the null model, 10 000 re-samplings were conducted.

The SKAT-O test was chosen due to its optimisation parameters to test each gene set's genotypes under the null model. This was shown to be particularly useful in exome sequencing studies due to its ability to decrease type I error [202]. The test was conducted under the small-sample size kurtosis adjustment, implemented by the SKAT package, as the sample size was under 2 000 individuals. Optimisation of the test statistics were modulated by the program and tested across

the default distribution of ρ (0, 0.1^2, 0.2^2, 0.3^2, 0.4^2, 0.5^2, 0.5, and 1) to minimise the p-value produced during analysis, which was then used as the test p-value. The results were corrected for multiple testing using a Bonferroni-corrected significance cut-off value of 0.05. This cut-off limit was chosen because of the small sample size available for the analysis. The SKAT-O procedure limited the number of tests conducted due to the collapsing of SNPs into gene regions, thus contributing to the retention of study power.

3.2.6 SKAT Common Rare Analysis

To test the combined effect association of both common and rare SNPs in the analysis, the common/rare SNP association test as part of the SKAT R package was used [179, 205]. The SNP filtration and prioritisation procedures followed the same procedure as SKAT-O testing, with the exception of MAF filtration to retain common SNPs. The SKAT input Set.ID, SSD and gene-set files were created as before and all duplicates and missing information was removed. A total of 13 270 gene-sets with 53 239 SNPs were analysed.

Null model creation was done as described for SKAT-O testing (Section 3.2.5).

The test performed made use of the SKAT Common Rare function [181] in the SKAT R package [179, 205]. This procedure involved testing common and rare variants in two groups separated based on an adaptive MAF value as a function of the total sample size. The p-values of the two groups were then combined and weights assigned according to MAF on a per SNP basis. Test results were then corrected for multiple testing as in SKAT-O testing at a p-value significance cut-off of 0.05.

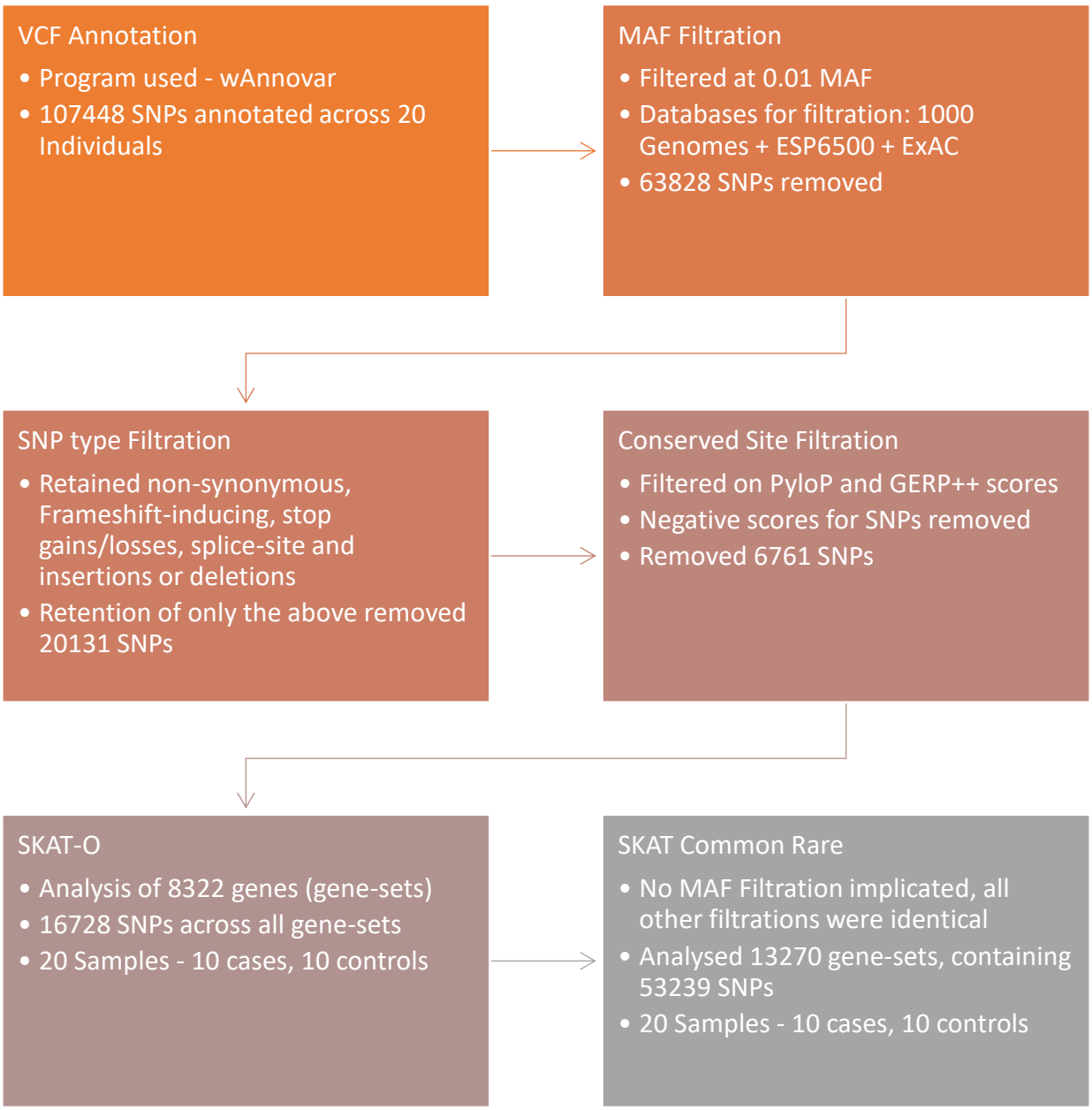

Figure 3.2: Overview of the SNP prioritisation procedures followed and association analyses used in the exome sequencing arm of the study. SKAT = Sequence Kernel Association Test.

3.2.7 Ingenuity Pathway Analysis (IPA) of SKAT results

Functional assignment and pathway analysis of the association results was necessary to prioritise genes of interest from the SKAT Common-Rare and SKAT-O analyses. Association results were

filtered for the retention of unadjusted p-values below a threshold of 0.01, this allowed for sufficient genes for input. Unadjusted p-values were used in downstream analyses as multiple testing correction led to no genes retaining significance. This may be due to the small sample size of the study and therefore a lack of power to detect rare associations. As unadjusted p-values were used, the likelihood of the included genes being false positives is high.

Using IPA core analysis, the top hits from both SKAT association tests were assessed independently [206]. Assessments conducted include overrepresentation analysis of pathways assigned to the input genes, network construction between gene products and genes, functional assignment, disease associations and assessment of shared regulatory molecules of the input genes. Analysis filtration parameters were defined to draw information from the Ingenuity Knowledge Base (genes only). Network construction was not constrained as exploration into related genes was desired, thus direct and indirect relationships were used. Filtration parameters for results pertaining only to *Homo sapiens* were applied to exclude any gene functions annotated through homology from other species. Additionally, filtration was applied to include only experimentally confirmed interactions, thus providing maximum confidence in the results. Using this information, genes of interest were prioritised according to pathways and functions thought to be involved in TBM, such as BBB regulation and genes displaying immunological functions. Network construction involved both direct and indirect interactions to connect input genes to one another. Direct interactions biased construction towards protein-protein binding of two input gene products, whereas indirect interactions allowed for the connection of input genes that function in the same pathway but do not necessarily interact with each other.

3.3 Results

The depth of the exome sequencing obtained in this study ranged from 35x to 82x. The mean depth across all 20 individuals was 52x, where coverage between 40x - 60x is generally considered adequate for exome sequencing [207]. A total of 107 448 SNPs were called across the 20 study individuals. SNP prioritisation for the SKAT-O analysis removed 63 828 SNPs that were above a MAF of 1%. An additional 20 131 SNPs were removed during filtration for the inclusion of non-synonymous and frameshift inducing SNPs, stop gains/losses, splice-site SNPs and insertions and deletions. A further 6 761 SNPs were removed as a result of their conservation score, leaving a total of 16 728 SNPs for SKAT-O analysis. The SKAT Common Rare prioritisation removed 35 589 SNPs to prioritise non-synonymous and frameshift inducing SNPs, as previously mentioned. A further 18 620 were removed based upon their conservation score, leaving a total of 53 239 SNPs for SKAT Common Rare analysis.

3.3.1 SKAT-O

Following SNP prioritisation, a total of 17 191 SNPs were available for analysis. Following removal of SNPs with overlapping base-pair positions, SKAT-O analysis was performed on 8 322 gene-sets containing 16 728 SNPs using a dichotomous phenotype of patients with TBM compared to healthy controls. None of the gene-sets were found to be associated after correcting for multiple testing using a Bonferroni significance level of $p = 0.05$. The gene-set with the lowest p-value was that of Zinc Finger Homeobox-3 (*ZFHX3*), $p_{unadjusted} = 4.63 \times 10^{-4}$. See Appendix 1 for gene-sets with $p_{unadj} < 0.01$.

3.3.2 SKAT Common Rare

SKAT Common Rare analysis testing was done to include both common and rare SNPs in our analysis and weighted as described in Chapter 2. This allowed for the inclusion of more gene-sets and SNPs, with 13 270 gene-sets and 53 239 SNPs analysed. A single significant association was identified by comparing TBM cases and controls and after correcting for multiple testing. The Centriolar Coiled-Coil Protein 110kDa (*CCP110*) gene was found to be associated with TBM (p

= $5.89x10^{-6}$) after correcting for multiple testing, with the combined effects of 1 rare SNP and 2 common SNPs contributing to the association. See Appendix 2 for gene-sets with $p_{unadj} = 0.01$.

3.3.3 Ingenuity Pathway Analysis® (IPA)

Functional pathway analysis allows for the interpretation of association study data results in a biological context. It is also useful in hypothesis generation and exploratory analysis of selected genes. This will allow future studies to target genes and pathways of interest rather than approaching the disease from a genome-wide context which proves to be less economical. IPA analysis was performed to explore the pathways affected and the associated functional annotations of each gene of interest from both SKAT analyses.

3.3.3.1 SKAT-O

A p-value cut-off of $p < 0.01$ was used to identify top hits for exploratory analysis using IPA core [206]. Pathway over-representation in the top hits of an association analysis can point towards the defective function of a particular pathway and can indicate possible involvement in disease pathogenesis. IPA provided a measure of overlap between the SKAT top hits and curated canonical pathways in the database. In the SKAT-O analysis, the pathways with the greatest proportion of overlap with the SKAT top hits are listed in Table 3.1.

Table 3.1: Highlighted canonical pathways in the SKAT-O analysis.

Pathway Name	p-value	Overlap between input and total genes in pathway (Ratio)	Overlapping Genes
Transcriptional regulation network in embryonic stem cells	$2.90x10^{-3}$	5%, (2/40)	*RIF1, ZFHX3*
α-tocopherol Degradation	$7.98x10^{-3}$	25% (1/4)	*CYP4F2*
Histamine Degradation	$2.57x10^{-2}$	7.7% (1/13)	*ALDH3B1*
Oxidative Ethanol Degradation III	$2.96x10^{-2}$	6.7% (1/15)	*ALDH3B1*
Fatty Acid α-oxidation	$3.15x10^{-2}$	6.2% (1/16)	*ALDH3B1*

Figure 3.3 shows the results for α-tocopherol degradation wherein both the p-value and overlap ratio remain significant as both are above the threshold. The -log p-values are calculated based upon the role that each gene plays in the pathway. Ratios are determined by the number of overlapping genes between the input genes and the total genes in the curated pathway. This pathway had a -log p-value of 2.098 with an overlap ratio of 25% as the input gene, cytochrome P450 family 4 subfamily F member 2 (*CYP4F2*), was 1 of 4 genes implicated in the process.

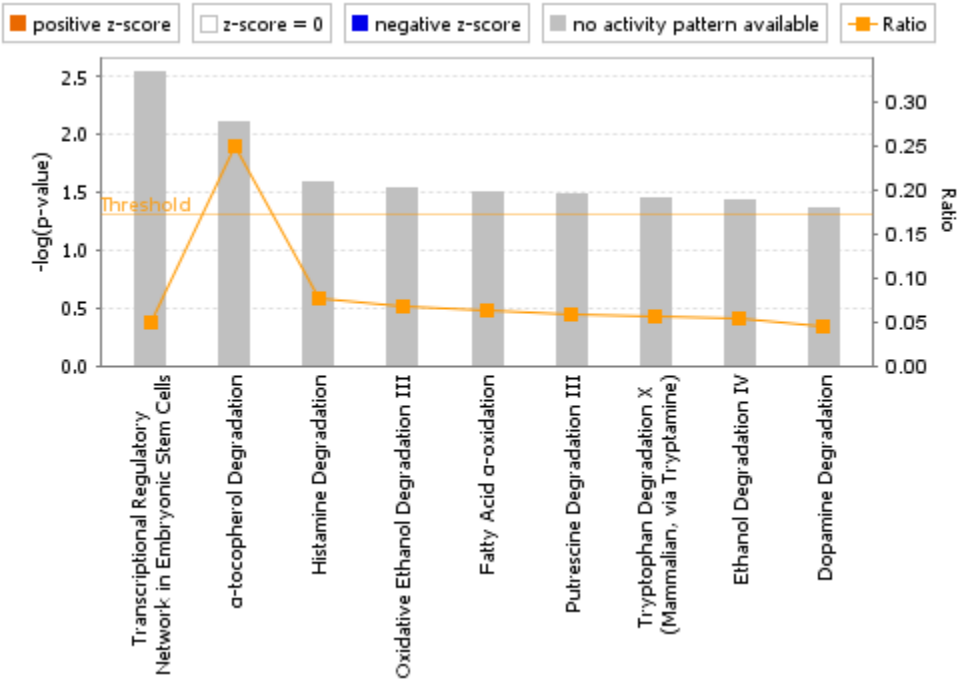

Figure 3.3: Canonical pathways highlighted in IPA analysis.

Genes that perform similar functions may provide insight into functional defects that influence TBM susceptibility. These functions may provide novel information into possible susceptibility pathways that influence disease pathogenesis due to a lack of redundancy measures in the host. Of interest were functional annotations that were thought to play a role in TBM or pTB pathogenesis, such as immune regulation, apoptosis or autophagy related gene functions. The top functional annotations are shown in Table 3.2. The Polyhomeotic homolog-1 (*PHC1*) was not considered as a viable candidate for further investigation as it functions in transcriptional regulation in embryonic stem cells and is associated with their differentiation rather than bacterial clearance [208]. As *CYP4F2* had been implicated in enriched pathways (Table 3.2) it was interesting that it was also implicated in several functional enrichment categories shown in Table 3.2. Investigations into the

role of the other genes listed in Table 3.2 did not reveal any plausible links with known TBM pathogenesis pathways or functions.

Table 3.2: Enrichments for molecular function in the SKAT-O analysis.

Functional Annotation	p-value range	Annotated Genes
Lipid Metabolism	$3.35x10^{-2}$ – $2x10^{-3}$	*CYP4F2, PIGG, GCKR*
Small Molecule Biochemistry	$3.35x10^{-2}$ – $2x10^{-3}$	*CYP4F2, ALDH3B1, PIGG, GCKR, ITPR3, ESRRA*
Vitamin and Mineral Metabolism	$3.15x10^{-2}$ – $2x10^{-3}$	*CYP4F2, ITPR3*
Cell Death and Survival	$4x10^{-3}$	*PHC1*
Cellular Assembly and Organisation	$2.77x10^{-2}$ – $4x10^{-3}$	*FARP2, TOPBP1, ESPL1*

Physiological system development is crucial to the overall functioning and performance of the system as a whole. Defects in system physiology may affect its ability to function as intended and the system may be vulnerable to pathogenic invasion. Table 3.3 shows that nervous system development was the most significantly affected physiological system when the input genes from the SKAT analysis were considered.

Table 3.3: The development and function of physiological systems most affected by genes highlighted by the SKAT-O analysis.

Physiological System	p-value range	Annotated Genes
Nervous System development	$4x10^{-3}$	*FARP2*
Embryonic Development	$3.15x10^{-2}$ – $4.85x10^{-3}$	*CC2D1A, NOD2*
Hair and Skin Development	$9.96x10^{-3}$ – $4.85x10^{-3}$	*CC2D1A, NOD2*
Renal and Urological Development	$9.96x10^{-3}$ – $5.22x10^{-3}$	*CC2D1A, NOD2*
Skeletal and Muscular Development	$4.89x10^{-2}$	*ZFHX3*

The gene implicated in CNS development (FERM, ARH/RhoGEF and Pleckstrin Domain Protein 2 (*FARP2*)) was found to be implicated in the control of neurite remodelling in the cerebellum. It

is unlikely that *FARP2* plays a role in TBM pathogenesis as neurite remodelling is likely to affect neural connections and their development [209].

The gene Nucleotide binding and oligomerisation domain 2 (*NOD2*) functions in bacterial recognition and autophagy, which are both pathways of interest in TBM. Table 3.3 shows that *NOD2* was highlighted across several developmental processes, but also functions in pathogen recognition in the immune system [210].

3.3.3.2 SKAT Common Rare

SKAT Common Rare was used to assess the cumulative effects of both common and rare SNPs with susceptibility to TBM. As in the SKAT-O IPA results, top hits from the SKAT Common Rare test ($p < 0.01$) were analysed using IPA® Core Analysis [206]. Interestingly, *CCP110* was not identified in any of the associated pathways, suggesting a lack of involvement. Pathways that were common to multiple genes in the SKAT Common Rare results are listed in Table 3.4. Significance was only retained for S-methyl-5'-thioadenosine Degradation II in terms of both a significant proportion of overlap of input genes with the total genes in the curated pathway and -log p-value. This is most likely due to the pathway only having a single gene ascribed to it, methylthioadenosine phosphorylase (*MTAP*), thus its inclusion is overrepresented in the analysis due to the lack of other contributing genes.

Table 3.4: Canonical pathways highlighted in IPA analysis.

Pathway Name	p-value	Overlap of input genes with known pathways	Genes Associated
S-methyl-5'-thioadenosine Degradation II	7.53×10^{-3}	100%, (1/1)	*MTAP*
Glycine Biosynthesis III	1.50×10^{-2}	50%, (1/2)	*AGXT2*
Role of Oct4 in Mammalian Embryonic Stem Cell Pluripotency	4.51×10^{-2}	4.4 %, (2/45)	*FAM208A, NR5A1*
Phototransduction Pathway	5.65×10^{-2}	3.9 %, (2/51)	*GNAT2, GUCA1C*
FXR/RXR Activation	6.85×10^{-2}	2.4 %, (3/125)	*APOF, G6PC, G6PC3*

Regulation of genes that function in similar pathways or perform similar functions may involve common regulatory proteins such as transcription factors. This may lead to enrichment of the dataset in certain regulatory factors. This is an important consideration as the basis for transcription factor function relies on the conservation of the recognition motif or sequence of the target protein. Modifications to this binding sequence would likely affect transcription factor target affinity and binding efficacy, thereby affecting control of the target gene. Several genes were found to be under the control of common regulatory elements, listed in Table 3.5. *CD44* was found to target 4 genes in the input dataset, namely DNA (cytosine-5) methyltransferase 1 (*DNMT1*), interferon induced transmembrane protein 2 (*IFITM2*), interleukin 1 receptor associated kinase 3 (*IRAK3*) and programmed cell death 4 (*PDCD4*). Both *IFITM2* and *IRAK3* play important roles in immune signal transduction pathways, whereas *DNMT1* functions in the regulation of gene expression and *PDCD4* is a tumour suppressor gene.

Table 3.5: Common regulatory elements for the top-hits of SKAT Common/Rare.

Regulatory Element	p-value
CD44	$3.10x10^{-4}$
PSMD4	$5.11x10^{-4}$
PPP2CB	$7.22x10^{-3}$
TIPIN	$7.22x10^{-3}$
TAL1	$1.21x10^{-2}$

Pathway analysis revealed that the most enriched disease annotation was that of cancer, with a p-value between $4.7x10^{-02}$ – $9.91x10^{-04}$, corresponding with 132 genes highlighted in the SKAT Common Rare analysis. Disease enrichments for neurological or pulmonary diseases were not found in the IPA analysis. Annotations for enriched molecular and cellular functions are shown in Table 3.6.

Table 3.6: Enrichments for molecular and cellular functions.

Functional Annotation	p-value	Top-Hit Genes Associated
Cellular Movement	$4.43x10^{-2}$ - $1.68x10^{-4}$	*A2M, GNRH1, MYO7A*
Cell Morphology	$4.43x10^{-2}$ - $4.42x10^{-3}$	*CYFIP1, MACF1, ROR1, ZFYVE16*
Cellular Assembly and Organisation	$4.43x10^{-2}$ - $4.42x10^{-3}$	*CYFIP1, MACF1, ROR1, KRT20, MAP2, PACS2, EXO5, ANLN, DDX11, ZFYVE16, SPAG16, NLGN4X, MYO7A*
Cellular Function and Maintenance	$4.43x10^{-2}$ - $4.42x10^{-3}$	*CYFIP1, MACF1, ROR1, KRT20, A2M, SPAG16, MAP2, NLGN4X, MYO7A*
Cell Cycle	$2.98x10^{-2}$ - $7.53x10^{-3}$	*PNTP1, HAVCR1, EXO5, CLSPN*

Enrichment analysis pertaining to genes functioning in physiological system development highlighted genes which function in neurological development as some of the top hits ascribed to the input gene list from the association analysis, as shown in Table 3.7.

Table 3.7: Enrichment of genes that correspond to the development and function of physiological systems.

Physiological system development and function	p-value	Genes ascribed to pathway
Nervous System	$3.71x10^{-2}$ - $1.68x10^{-4}$	*A2M, CX3CR1, CYFIP1, GNRH1, MACF1, NLGN4X*
Embryonic Development	$3.71x10^{-2}$ - $7.53x10^{-3}$	*GNRH1, NR5A1, NLGN4X*
Hair and Skin	$7.53x10^{-3}$	*KRT20*
Haematological system	$4.43x10^{-2}$ - $7.53x10^{-3}$	*CX3CR1, MARCO, A2M, PNPT1, HAVCR1, CLC*
Humoral Immune Response	$7.53x10^{-3}$	*HAVCR1*

This result was relevant given that hits for the enrichment of genes involved in the proper development and functioning of neurological, embryonic, haematological and humoral immune response systems were all highlighted. This could point towards defects in the development and function of several critical physiological functions forming part of the aetiology of TBM.

IPA generates networks during core analysis between input genes and populates the relationships between input genes with known interactors from the Ingenuity Knowledge Base. Network analysis showed that two of the input genes, receptor tyrosine kinase-like orphan receptor 1 (*ROR1*) and *DNMT1*, both regulate cadherin-1 (*CDH1*), a critical tight junction protein in the BBB (Figure 3.4). *ROR1* was also identified in a number of functional enrichments (Table 3.6), including cell morphology, cellular assembly and organisation and cellular function and maintenance. *ROR1* has also been shown to regulate other critical tight junction proteins at the BBB through the use of the transcription factor Snail-1 (*SNAI1*) [211].

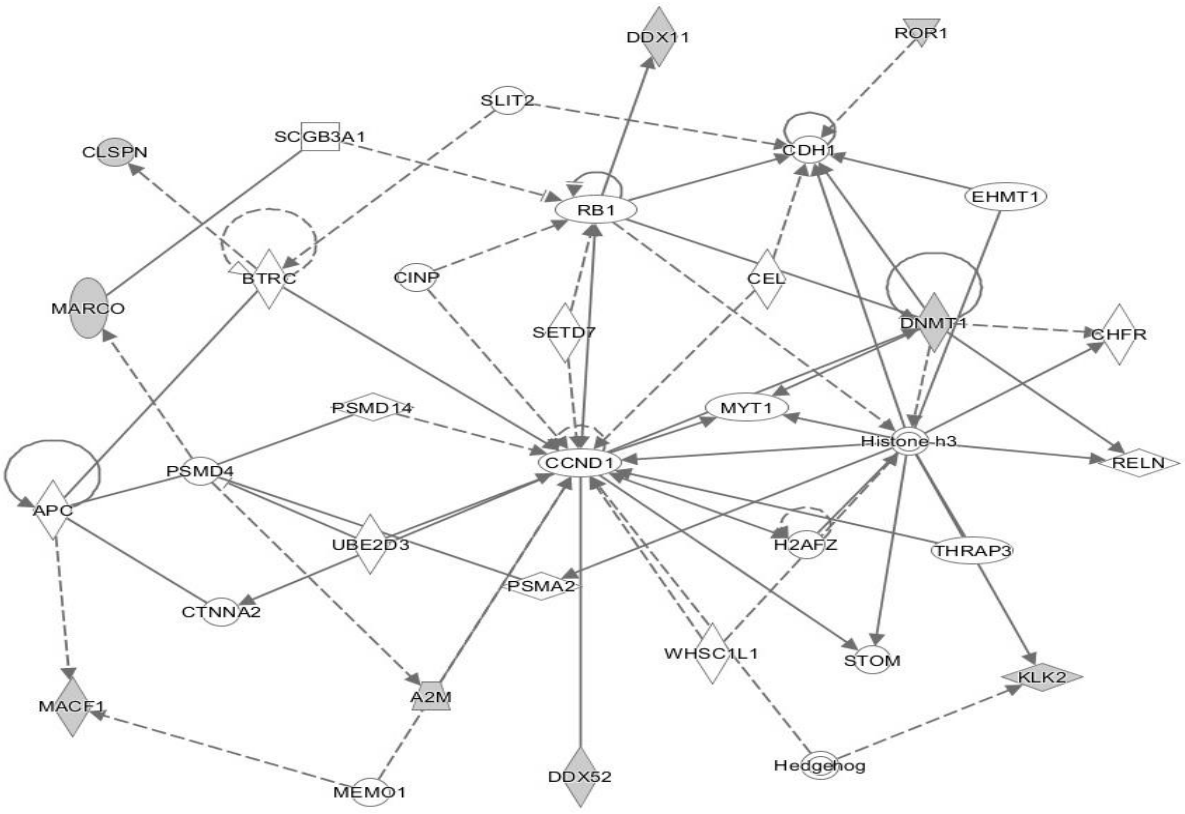

Figure 3.4: The relationships between CDH1, ROR1 and DMNT1. ROR1 interacts indirectly through other proteins to exact its regulatory function upon CDH1, denoted by the broken line. DNMT1 interacts with CDH1 in a direct protein-protein interaction, denoted by the solid line.

DNMT1 functions in gene regulation through the process of methylation of CpG sites and is a key epigenetic regulator during development. It exerts its methylation activity upon the promoter region of *CDH1* specifically in order to lower its expression [212].

Finally, the top hit genes were analysed for possible involvement in pathways that have been previously implicated in TBM or pTB pathogenesis, for example immune system regulation and autophagy related processes. As shown in Table 3.8, an additional 12 genes were found that may play an important roles in known TBM pathogenesis pathways, but have yet to be investigated in the context of this disease.

Table 3.8: Genes from the SKAT Common/Rare top hits that function in pathways thought to be involved in TBM pathogenesis.

Gene ID	Gene Name	Function and Pathway
A2M	Alpha 2 Macroglobulin	Cytokine transporter, IL-6 pathway
ATG10	Autophagy-related 10	Autophagy – Autophagosome formation
C4BPA	Complement component 4 binding protein, alpha	Regulation of the complement system
CX3CR1	Chemokine (C-X3-C motif) receptor 1	Leukocyte adhesion and migration, activation of microglia
IRAK3	Interleukin 1 receptor associated kinase 3	Component of the toll/IL-1 pathways, negative regulator of TLR signalling
MARCO	Macrophage receptor with collagenous structure	Innate immune signalling, Recognition receptor
IFITM2	Interferon induced transmembrane protein 2	Restriction of *M.tb* infection, Phagosome maturation

3.4 Discussion

3.4.1 Exome Sequencing: SKAT-O

Assessment of the proportion of rare variants was performed using the SKAT-O gene-set association test. Results indicted a lack of significant association following adjustment for multiple testing correction and using gender as a single covariate.

Pathway enrichment analyses were done using IPA to determine possible susceptibility pathways in which multiple genes may function. In doing so, *CYP4F2* (Table 3.2), was highlighted as functioning in the α-tocopherol degradation pathway ($p_{enrichment} = 1x10^{-3}$). Further functional investigation of *CYP4F2* showed involvement in the early stages of leukotriene B4 (LTB4) degradation [206, 213, 214]. This is of interest as the leukotriene A4 hydroxylase (*LTA4H*) locus has previously been shown to play a role in in TBM susceptibility due to its effects on TNF-α levels [98, 99]. Specifically, the control of TNF-α concentrations was found to differ based on genotype. Wild-type homozygotes showed hyper-inflammation due to an excess of LTB4, which in turn led to an excessive TNF-α response. The involvement of *CYP4F2* in LTB4 regulation is mediated by degradation of LTB4 through ω-hydroxylation leading to the attenuation of TNF-α signalling thus modulating the inflammatory response along with the LTA4H enzyme [214]. In addition, *CYP4F2* and *CYP4F3* have been shown to metabolise lipoxin A4 (LXA_4) and lipoxin B4 (LXB_4), both known anti-inflammatory lipoxins [215]. An excess of LXA_4 in particular has been shown to result in a hypo-inflammatory phenotype due to its anti-inflammatory actions on *TNF*-α. Therefore a contributing factor to this excessive LTB_4 and LXA_4 could be attributed to non-synonymous SNPs in *CYP4F2* causing a loss of protein function, with rs3952537 predicted to be damaging by SiFT, Polyphen-2 and loss-of-function protein effect predictions as provided by the variant effect predictor (VEP) accessed through Ensembl [216]. In fact two *CYP4F2* SNPs, rs1272 and rs3093158, have been implicated in Crohn's Disease susceptibility in Canadian children [217]. To fully determine the role that *CYP4F2* may play in the regulation of TNF-α concentration, the interplay between these molecules needs to be functionally assessed *in vivo*, particularly in TBM patients.

Pathways highlighted by IPA may provide insight into TBM pathogenesis. These results can therefore be used to generate hypotheses. This approach aids in focussing future studies on candidate genes which function in relevant disease pathways rather than viewing the data from a

genome-wide context. Table 3.3 highlights genes from the IPA analysis that were of interest considering their roles in immune system regulation, autophagy and CNS development. Nucleotide binding and oligomerisation domain (*NOD2*) formerly known as *CARD15*, was found to be important in the development of three independent physiological systems including embryonic, hair and skin and renal developmental processes. *NOD2*, however, is of greatest interest for its role in innate immunity and is expressed on the surface of astrocytes and microglia in the CNS and functions as a recognition receptor for *S. pneumoniae* LPS derived muramyl dipeptide [218]. Additionally, it has been implicated in the generation of IL-6 and TNF-α pro-inflammatory cytokines through the stimulation of NFκB towards other forms of bacterial meningitis caused by *N. meningitidis* and *B. bergdorferi* [210]. Through the use of two murine models of $NOD2^{+/+}$ and $NOD2^{-/-}$, acute CNS inflammation was demonstrated with increased levels of CCL3, TNF-α and decreased levels of anti-inflammatory IL-10 in $NOD2^{+/+}$ mice [218]. This demonstrates a plausible role for *NOD2* in the generation of a detrimental inflammatory response as observed in TBM. *NOD2* may represent a viable candidate gene for investigation in the context of TBM considering that the recognition of peptidoglycan derivatives would also be possible given the presence of peptidoglycan in the cell walls of *M.tb* [219]. Genotypic association studies in our study population of the p.Arg702Trp, p.Gly908Arg and p.Phe1007Ser polymorphisms which have been previously associated with Crohn's disease, showed a lack of association with pTB susceptibility [220]. These associations, however, have not been examined in the context of TBM patients to determine susceptibility.

The lack of association found in the SKAT-O analysis and the lack of representation of the top-hit genes in TBM pathogenesis pathways is not surprising when the frequency of the SNPs used is considered. Only a small number of SNPs were included in the analysis leading to lowered study power to detect associations with any gene. Additionally, due to the rarity of the SNPs, a further decrease in power will have been observed as a result of the small sample size of 20 individuals investigated. These limitations may have increased the false positive rate of the results found in the SKAT-O analysis and thus candidate genes identified should be individually genotyped using tag-SNPs in a much larger TBM cohort to glean a better understanding of the role of each candidate gene in TBM pathogenesis, assuming that associations are found between individual SNPs and TBM susceptibility [221].

3.4.2 Exome Sequencing: SKAT Common Rare

In contrast to the lack of significance found in the SKAT-O analysis, the SKAT Common Rare analysis found a single gene-set after correction for gender and multiple testing. *CCP110* was the only significant result ($p=5.89x10^{-6}$). *CCP110* functions in cell-cycle maintenance, centrosomal duplication and is a negative regulator of ciliogenesis [222]. Given its functional annotations and its role in centrosomal duplication, it seems unlikely that *CCP110* will have a direct effect on TBM pathogenesis pathways. Considering that proper correction for all covariates was not possible, it is plausible that *CCP110* is a false positive. *CCP110* was also not highlighted in any of the pathways or functional annotations during IPA analysis. However, since none of the covariates, besides age, were found to be significantly different between the case and control groups it is also possible that *CCP110* has an unknown function that may play a role in TBM.

The proper development of physiological systems such as the nervous system is critical especially when the age of onset of TBM is considered [223]. The IPA analysis of the SKAT Common Rare top associations revealed 6 genes known to function in the development of the nervous system. This is critical in TBM pathogenesis as incomplete or poor development of protection mechanisms in crucial nervous system barriers such as the BBB, may expose the nervous system to pathogenic attack. One of these genes, alpha 2 macroglobulin (*A2M*), functions as a protease inhibitor for all four classes of proteases and is also a cytokine transporter for TNF-α, IL-6 and IL-1β amongst others [224]. *A2M* is regulated by *IL-6*, a prominent cytokine in TB pathogenesis, through signal transduction using *STAT3* [225]. *A2M* along with IL-6 and C-Reactive-Protein (CRP) levels in the CSF have been shown to be indicative of blood-cerebrospinal fluid barrier (BCB) damage in bacterial meningitis as part of the acute phase reaction [226]. The acute phase reaction involves the rapid activation of the inflammatory response by both the innate and adaptive immune systems to pathogenic challenge [227]. *A2M* has been reported as a possible biomarker for disease progression in both pTB and all forms of EPTB largely due to its role in IL-6 transport [228, 229]. Its inhibitory properties mediated by protease activities have shown a distinct role in the regulation of matrix metalloproteases (MMPs, specifically), most notably *MMP-9* and *MMP-2*, in addition to which it regulates transforming growth factor beta (*TGF-β1*) [230, 231]. Notably, MMP dysregulation may contribute greatly towards nervous system damage observed in TBM as MMPs damage BBB components and are a leading cause of intracranial haemorrhage [231, 232]. *TGF-β* has been implicated in other forms of bacterial meningitis as a suppressor of polymorphonuclear

leukocytes which release reactive oxygen species and IL-1β amongst others [233]. These functions do assist in bacterial clearance, but also contribute to the inflammatory damage to the nervous system. *TGF-β* deletions were found to increase the recruitment of polymorphonuclear leukocytes to the CNS and ultimately enhanced bacterial clearance and significantly decreased vasculitis. In light of this, it could be suggested that *A2M*-mediated clearance of TGF-β could induce similar effects on bacterial clearance and the occurrence of vasculitis with concurrent increases in inflammation. Due to its roles in the degradation and transport of MMP, TGF-β and IL-6, *A2M* is a good candidate gene for future studies as a potential biomarker and for functional studies in the context of TBM.

A second gene highlighted in the category of nervous system function and development was *CX3CR1*. Its ligand, CX3CL1 or Fractalkine, functions in leukocyte adhesion to endothelial cells and the recruitment of T and B lymphocytes [234, 235]. *CX3CR1* encodes its receptor and plays a crucial role in signal transduction leading to the activation of microglia [236]. Its developmental functions may be restricted to microglial activation in terms of neuronal pruning during the basal development of the brain, but it has also been implicated in angiogenesis [236]. Under pathogenic challenge, microglia are activated through CX3CL1-CX3CR1 binding and production of TNF-α, IL-1β and IL-6 is induced [237]. Expression of these cytokines leads to an inflammatory response which has been shown to induce ischemic injury if cytokine expression is not ablated [237]. Studies involving *CX3CR1*$^{-/-}$ mice found that knockout mice exhibited reduced immune-cell recruitment to the site of infection, however, mycobacterial clearance was not significantly affected [238]. This suggests the presence of redundancy measures with regards to mycobacterial clearance, where ablation of *CX3CR1* expression may influence the amount of ischemic injury observed in TBM patients, but this requires further investigation. CX3CR1 protein expression is upregulated in the presence of TGF-β leading to the inhibition of fractalkine-mediated signalling and a resultant decrease in microglial activation [239]. Through this mechanism, the microglial activation response towards *M.tb* may be affected, especially when the number of immune-related genes regulated by TGF-β is considered. Given the roles *CX3CR1* plays in microglial activation and inflammatory responses, further functional studies should be conducted to further our understanding of *CX3CR1* involvement in a TBM setting.

Additional genes ascribed to nervous system development functions include: *CYFIP1*, *GNRH1*, *MACF1* and *NLGN4X*. These 4 genes display strictly developmental functions with particular focus on neuronal body or axonal extension. Cytoplasmic FMR1-interacting protein 1 (*CYFIP1*) is a repressor of fragile X mental retardation 1 (*FMR1*), which functions in female reproduction. Additionally, *CYFIP1* functions in axonal outgrowth [240]. Gonadotropin-releasing hormone-1 (*GNRH1*) displays critical functionality in reproduction through the release of gonadotropins. It is also responsible for the migration of *GNRH1*-expressing neurons from the developing nasal passages to the forebrain during early development [241]. Neuroligin 4, X-linked (*NLGN4X*) is a neuronal cell surface protein that binds specifically to presynaptic neurexins to induce synaptic remodelling during brain development [242]. Microtubule-actin crosslinking factor 1 (*MACF1*) enables the interaction between actin and microtubule subunits at the cell periphery and assists in neuronal development as a positive regulator of the Wnt pathway which is responsible for tissue generation and axonal guidance [243]. These genes may not necessarily play a direct role in TBM pathogenesis, but their developmental role leads to the question of whether a developmental aspect could contribute to TBM pathogenesis through the introduction of vulnerable tissues. These tissues may display "leaky" characteristics making them vulnerable to bacterial migration allowing easier migration into the CNS in carriers with SNPs in the genes listed above.

Investigation of the haematological system development functional annotation, highlighted both *CX3CR1* and *A2M*, which have been discussed above, as well as macrophage receptor with collagenous structure (*MARCO*). MARCO is a member of the class A scavenger receptors and functions in the recognition of mycobacterial cell surface trehalose dimycolate (TDM) moieties through an extracellular C-terminal, scavenger receptor cysteine-rich domain (SRCR) [244]. This recognition is mediated through interactions with both TLR2 and CD14 and induces NF-κB activation and the production of TNF-α, IL-6 and IL-12 [245]. As a result of the recognition of TDM, marked increases in the cell surface expression of MARCO have been observed in response to BCG infection in mice [246]. SNPs have also been investigated in *MARCO* for association with susceptibility to pTB in both the Chinese Han and the Gambian populations [247, 248]. A SNP, rs17009726, was associated with increased susceptibility to pTB ($p = 9.27x10^{-5}$ [OR = 1.65, 95% CI: 1.35-2.05]) [247]. A haplotype block GC containing rs17009726, p = 0.0001 was associated with increased susceptibility to pTB. A second haplotype block TGCC including rs17795618, rs1371562, rs6761637 and rs2011839, was also associated with increased susceptibility to pTB

with p = 0.029 [247]. Three SNPs: rs4491733, rs12998782 and rs13389814 were found to be associated with increased susceptibility to pTB in the Gambian population [248]. Additionally a single SNP was found to increase resistance to pTB, rs7559955[248]. MARCO cell surface receptors expressed on glial cells also have a proven role in meningitis defence which has been investigated in the context of *S. pneumoniae* and *N. meningitidis*, the two most common causes of bacterial meningitis [245]. A 5-fold (±0.8-fold) increase in MARCO protein expression was observed in astrocytes following a 12 hour infection with *N. meningitidis* [245]. Cytokine production was also shown to increase as measured by IL-1β production. Increased cell surface expression and cytokine production was significant for both astrocytes and meningeal cells [245]. Although *MARCO* has proven action as a recognition receptor for *M.tb* specific surface moieties in pTB and recognition of other forms of bacterial meningitis, no study has investigated *MARCO* polymorphisms in a TBM setting. Considering its involvement in both disease types is a necessity to better understand the pathogenesis of TBM.

The construction of networks between input genes provided an overview of the interactions between the genes of interest and gives a view of the regulatory effects of each interaction. The analysis of one of these networks (Figure 3.4) highlighted the relationship between *ROR1* and *CDH1*. *ROR1* is a modulator of neurite growth in nervous system development and a mediator of endothelial-mesenchymal transition (EMT) during development [211, 249]. EMT is the process during which endothelial cells undergo phenotypic transition. This is initiated through the loss of classic endothelial cell markers, one of these being vascular endothelial cadherin (*CDH5*). These transitioning cells display a mesenchymal phenotype and begin to express cell products specific to mesenchymal cells. This transition enables increased motility and migration into neighbouring tissues and this process can be induced through TGF-β [250]. *ROR1* was found to regulate this process and inhibitory studies have shown a concurrent decrease in the expression of proteins linked with EMT such as vimentin, SNAIL-1 and -2 and zinc finger E-box binding homeobox 1 (ZEB1). The expression of *CDH1* was increased as was the expression of several tight junction proteins [211]. This shows that *ROR1* plays a regulatory role in the expression of tight junction proteins during embryogenesis, as it is expressed at low levels in adult tissues [251]. The regulation of tight junctions is mediated by the Snail-1 and -2 transcription factors as increased signalling through the binding of the WNT5A ligand to the ROR1 protein stimulates the production of the T-cell Factor (TCF) and lymphoid enhancer factor (LEF) transcription factors which bind to the

promoter of *SNAI1* [252]. Snail-1 and Slug (*SNAI2*) have been shown to regulate tight junction proteins zona occludens-1 (TJP-1), occludin (*OCLN*) and VE cadherin (*CDH5*) in bacterial meningitis caused by group B *Streptococcus* (GBS) [253]. This downregulation induced BBB disruption as permeability was increased. This was modelled in zebrafish which exhibited significantly increased bacterial loads and mortality compared to non-infected controls [253]. The induction of *SNAI1* was found to be dependent upon extracellular signal-related kinases 1/2 and mitogen activated protein kinase (ERK1/2/MAPK) signalling cascades mediated through the recognition of cell wall components by the TLR pathway. This mechanism of *SNAI1* BBB disruption was further supported through the investigation into the mechanics surrounding an observable downregulation of claudins-7 and -10 in both *S. pneumoniae* and *H. influenzae* infection [254]. Additionally, *SNAI1* induction was preceded by MAPK and TGF-β signalling pathways which were initially stimulated through TLR pathogen recognition.

The functional role of *ROR1* has yet to be properly elucidated and studies into its functional mechanics are sparse. Therefore implicating *ROR1* as a candidate gene for future study needs to be done with caution as our knowledge of its biological role and our knowledge of biological pathways as a whole is largely fragmented. Its ligand wingless homolog 5A (*WNT5A*), was shown to be differentially expressed in a microarray-based assessment of gene expression during *M.tb* infection [255]. Its induction was found to be TLR and NF-κB mediated, a novel pathway that has not been previously described, and expression was restricted to antigen-presenting cells such as macrophages. Binding of WNT5A to its other receptor frizzled-5 (FZD5) led to an increase in the IL-12 and IFN-γ production of peripheral-blood mononuclear cells following stimulation [255, 256]. Investigations into the concurrent activation of both TNF-α and WNT5A have led to the description of two subsets of cytokines that only partially overlap but are regulated by both in tandem, showing that WNT5A cytokine production is independent of TNF-α induction [255, 257]. WNT5A signalling has been implicated in the inflammatory phenotypes observed in other diseases such as psoriasis, metabolic dysfunction during obesity, sepsis and rheumatoid arthritis [257–263]. *WNT5A* expression was also observed in granulomas taken from *M.tb* patients [255]. *WNT5A* SNPs have been investigated for association with pTB and no significant associations were found [264]. These results, however, do not mean that SNPs in *WNT5A* are not associated with TBM. Given the upregulation of expression of *WNT5A* in macrophages and the consequent induction of pro-

inflammatory cytokines, studies into *ROR1* expression profiles in macrophages in response to mycobacteria should be assessed.

A single SNP, rs7527017, was analysed in *ROR1* and induces a non-synonymous amino acid change from a threonine to either a lysine or a methionine, depending on the transcript isoform of *ROR1*. rs7527017 has a MAF of 0.07695 (p.Thr518Met) in the African population according to the Exome Aggregation Consortium (ExAC) [194]. The threonine to lysine substitution was not observed in the African population. When analysed using VEP SNP effect prediction software, PolyPhen [265] and SiFT [266] effect prediction tools estimated damaging effects for the threonine to methionine substitution of 0.04 and 0.63, respectively. The threonine to methionine substitution induced by rs7527017 creates a new start codon in the middle of the transcript. Functional prediction using the loss-of-function tool (LoFtool) as a VEP plugin, showed a damaging score of 0.189, implying that the introduction of a new start codon would render both transcripts non-functional [216]. To date this SNP has only been studied in B-cell chronic lymphocytic leukaemia using a very small sample size of 15 individuals and was not found to be a distinguishing factor between disease subtypes [267]. Considering the downstream effects of *ROR1* signalling upon tight junction associated genes, the effects that WNT5A exerts upon pro-inflammatory cytokine release when bound to FZD5 and the implications of the creation of a transcription start codon created by rs7527017, *ROR1* should be considered for future genotyping and functional studies in larger cohorts.

Examination of the functional annotations ascribed to all genes in the top hits of the SKAT Common Rare analysis identified genes that function in pathways such as autophagy and other immune responses that may contribute towards TBM pathogenesis. One such gene identified was autophagy-related 10 (*ATG10*). Autophagy is a process whereby cytosolic components such as damaged organelles are phagocytosed and subjected to lysosomal degradation [268]. *ATG10* specifically functions as an E2-like enzyme that mediates ubiquitin modifications in phagosomal maturation and cargo recruitment [269]. The role of autophagy in TB is well established as a method for the persistence and evasion of killing mechanisms by *M.tb* [270]. Investigation into the effect of *ATG10* SNPs on gene function and consequent susceptibility to *M.tb* infection have been assessed [270]. No association was found for any of the SNPs studied, however, increased IL-8 production was observed in carriers of the rs1864183 SNP, $p = 0.04$. This result is interesting in

the context of TBM as IL-8 concentrations have been found to be significantly increased in the CSF prior to TBM treatment and have been suggested as a potential biomarker for differential diagnosis [271]. SNP genotyping and functional knock-out studies of the *ATG10*-mediated expression of IL-8 in a TBM cohort or model organisms may aid in determining the role of *ATG10* in TBM disease.

Three additional genes were identified in the SKAT Common Rare analysis. These 3 were mainly involved in innate immune system signalling pathways functioning in a regulatory capacity or as signal transducers. The first of these, interferon induced transmembrane protein-2 (*IFITM2*), has been shown to mediate antiviral responses in previous investigations [272–275]. Studies assessing its effects during *M.tb* infection led to an expansion of its functional properties in that a role in the restriction of *M.tb* infection was identified and described [276]. Through the use of *IFITM1-3* knockdown using RNA interference, significantly increased growth of *M.tb* was observed in both monocytic and alveolar cell types [276]. Individual overexpression assays were then conducted to determine the independent effect on *M.tb* survival in these cells and *IFITM2* displayed a 24.7% reduction in growth. Expression of *IFITM* genes was found to be mediated by both TLR2 and -4 receptor signalling upon *M.tb* recognition. *IFITM2* specific expression was observed through both the TLR receptor pathway and independently through the IL-6 and TNF-α pathways [276]. This demonstrates a redundancy measure for *IFITM2* signalling upon *M.tb* recognition suggesting an important role in *M.tb* containment. Further investigations have demonstrated the co-localisation of *IFITM3* with late *M.tb* phagosomes and the overexpression of *IFITM3* has been shown to contribute towards phagosomal acidification. In the case of *IFITM2*, co-localisations have not been assessed. However, given the redundancy measures resulting in its expression and its demonstrated role in *M.tb* containment it cannot be ruled out as possibly playing a role in TBM disease development warranting its inclusion as a candidate gene.

The complement system is a part of the innate immune system through which complement proteins enhance bacterial killing and is mediated by phagocytic cells [277]. Phagocytes are stimulated through complement protein cascades leading to the recruitment of additional phagocytic cells and increased immune inflammation [277]. The complement system is activated through three independent pathways namely; the classical, lectin and alternative pathways [277]. Complement component 4 binding protein alpha (*C4BPA*) functions in an inhibitory capacity upon the

complement pathway by preventing the assembly of the C3 convertase enzyme complex for the classical and lectin pathways (C4bC2a). This is coupled with the accelerated degradation of the C3 complex [278]. C4BPA proteins also serve as co-factors to serine protease factor I (FI) to induce C4 cleavage in either fluid phases or deposited upon cell surfaces, thereby inhibiting complement system activation through either the classical or lectin pathways [278]. Despite the host protective roles that C4BPA proteins exert, they may also be captured on the surface of pathogens thereby rendering pathogens resistant to the effects of the complement system [279]. Complement activation in the lung during *M.tb* infection is mainly induced through the classical complement pathway [280]. C4 proteins were highly expressed in pTB patients when compared to control individuals. This was coupled with concurrent expression of CRP which has been shown to target C4BPA proteins and assist in the modulation of complement inhibition [280, 281]. C4BPA proteins have also been linked with the pathogenesis of other forms of meningitis caused by *N. meningitidis* as activation is also mediated by the classical pathway [282]. C4BPA binding to the cell surface of *N. meningitidis* was observed and cell surface binding was enabled through binding to bacterial Porin A (PorA) subunits [282]. Critically, this binding enabled bacterial resistance to the complement system as C4BPA remained functionally active even in its bound state and was able to mediate the inhibition of the complement system [282]. Associations with protection against *N. meningitidis* infection have been found between Complement Factor H (*CFH*) ($p = 2.2x10^{-11}$ [OR = 0.64, 95% CI: 0.56-0.73]) and additionally with Complement Factor H Receptor 3 (*CFHR3*) ($p = 4.6x10^{-13}$ [OR = 0.63, 95% CI: 0.55-0.71]) [283]. Both CFH and CFHR3 proteins are inhibitors of complement system activation which act along with C4BPA to modulate the complement response [283]. This study demonstrates that SNPs in genes encoding inhibitory proteins of the complement system may contribute to host protection against disease as they are unable to be captured on the cell surface of pathogens thus allowing full host complement activation. CFH proteins have been shown to be captured on the cell surface of *N. meningitidis* to evade host complement-mediated phagocytosis [283]. Functional studies have shown a distinct link between *M.tb* and other forms of bacterial meningitis with the complement system and it represents a genuine method of immune subversion. The effects of this relationship still require assessment in the setting of TBM and SNPs in other complement system genes, particularly *C4BPA*, upon host susceptibility require investigation.

The recognition of microbial PAMPs is critical in mounting an efficient immune response against *M.tb* infection. One of these critical pathways of PAMP recognition is the TLR signalling pathway. TLR recognition of PAMP induces signal transduction through adapter proteins MyD88 and IRAK [284]. The *IRAK* gene family consists of *IRAK1* and *IRAK4*, both of which are active kinases and 2 inactive kinases *IRAK2* and *IRAK-M* (*IRAK3*) [285]. The expression of *IRAK-M* is restricted to macrophage and monocyte cell types, whereas the other IRAK proteins display ubiquitous expression [284]. *IRAK-M* is a negative regulator of the TLR signalling pathway by preventing the dissociation of the IRAK1-IRAK4 complex from the MyD88 adapter protein [284]. In doing so, this prevents the formation of IRAK-TRAF6 complexes which continue the signalling cascade leading to the production of pro-inflammatory cytokines such as IL-12 p40 and TNF-α [284, 286]. In *IRAK-M* $^{-/-}$ cells observable increases in resultant cytokine production following TLR stimulation was noted. This was coupled with increased inflammation which was detrimental to the host [287]. Studies into the specific role of *IRAK-M* in *M.tb* infection have shown a marked decrease in Th1 immune responses in the lung [288]. This decrease is mediated by the DNAX-activating protein 12kDa (DAP12) transcription factor which is required for *IRAK-M* expression, which in turn induces IL-10 production, an anti-inflammatory cytokine, leading to the immunosuppressed phenotype of antigen presenting cells [288]. This immunosuppressed phenotype was also observable through concurrent low-levels of nitric oxide synthase 2 (NOS-2) protein expression that aids in bacterial clearance [285]. The DAP12/IRAK-M/IL-10 axis therefore represents a susceptibility pathway in TB pathogenesis, the effects of which have not yet been elucidated in TBM infected patients. Considering the effects of TB with respect to the suppression of the Th1 response, mediated in part by *IRAK-M* expression, future investigations using SNP genotyping should be done to determine association of *IRAK-M* SNPs with pTB or TBM susceptibility. In addition, investigation into the tandem effects of SNPs in *IRAK-M*, *C4BPA* and *IFITM2* should be undertaken to determine their overall effect on innate immune system function in TBM patients.

In 2 of the analyses conducted in this study, secretion regulating guanine nucleotide exchange factor (*SERGEF*) was highlighted. Besides the functional annotation of being a guanine exchange factor, little is known about *SERGEF* function. *SERGEF* was identified by both SKAT analyses, while SKAT-O assigned an unadjusted p-value of $5x10^{-3}$ and tested a single rare SNP. SKAT Common Rare analysis showed a lower p-value of $4x10^{-4}$, likely attributable to 3 SNPs being tested

compared to a single SNP, rs377166311, which was the only shared SNP between analyses. The other two SNPs tested by SKAT Common Rare (rs10788 and rs1528) were both filtered out of the SKAT-O analysis due to their MAF being 0.13 according to the 1000 genomes, ESP6500 and ExAC databases [192–194].

The fact that both *SERGEF* was highlighted in multiple analyses increases the possibility that this gene, and in particular the SNPs attributed to the gene, may play a role in TBM pathogenesis. Assessment of the functional role of *SERGEF* in TBM still requires investigation. While it may be possible that SNPs found in the exome sequencing may tag SNPs that were found in the top hits of the GWAS or vice-versa, additional validation in a larger cohort of these results is necessary.

Genome-Wide Association Study Chapter 4

Chapter 4: Contents

Chapter 4: Genome-Wide Association Study

4.1 Introduction

Most sites in the genome are consistent across all humans, but approximately 10 million (1%) of all loci are polymorphic, accounting for the variation between individuals [289]. Differences in allele frequency have been observed between population groups as well. This variability in allele frequency accounts for inter-population differences and consequent variation observed in disease susceptibility between populations [290].

Genotyping arrays from Affymetrix and Illumina both utilise the hybridisation of fragmented single-stranded genomic DNA to array probes designed to interrogate a specific site sub-sequence (see Figure 4.1) [289]. Hybridisation signal intensity is then assessed on a per-probe basis to determine the affinity between the target and probe sequence. Processing and analysis of raw intensity readings often lead to genotyping accuracies of over 99.5%.

SNP selection for arrays is a major hurdle faced during design due to space constraints. Selection strategies are therefore geared towards the inclusion of SNPs previously known to be associated with particular diseases that share similar pathologies. The Illumina® Infinium Multi-Ethnic Genotyping Array (MEGA) (San Diego, CA, USA) was recently designed in conjunction with asthma studies of African ancestry individuals in the USA (CAAPA), the consortium of Population Architecture in Genomics and Epidemiology (PAGE) and from phase 3 of the 1000 genomes project. SNP selection included functionally relevant SNPs and those previously associated with asthma and other respiratory diseases such as TB. Amongst these, select genome-wide tag SNPs were selected to increase the power for detection of association and exome sequencing content from over 12 000 individuals. Given the necessity for replication of associations with complex diseases across population groups, SNP content from African-specific ancestries were included. In total the number of SNPs included was 1.7 million, making it the optimal choice for use in research in admixed populations. The admixture that is observed in the SAC is predominantly from African ancestry.

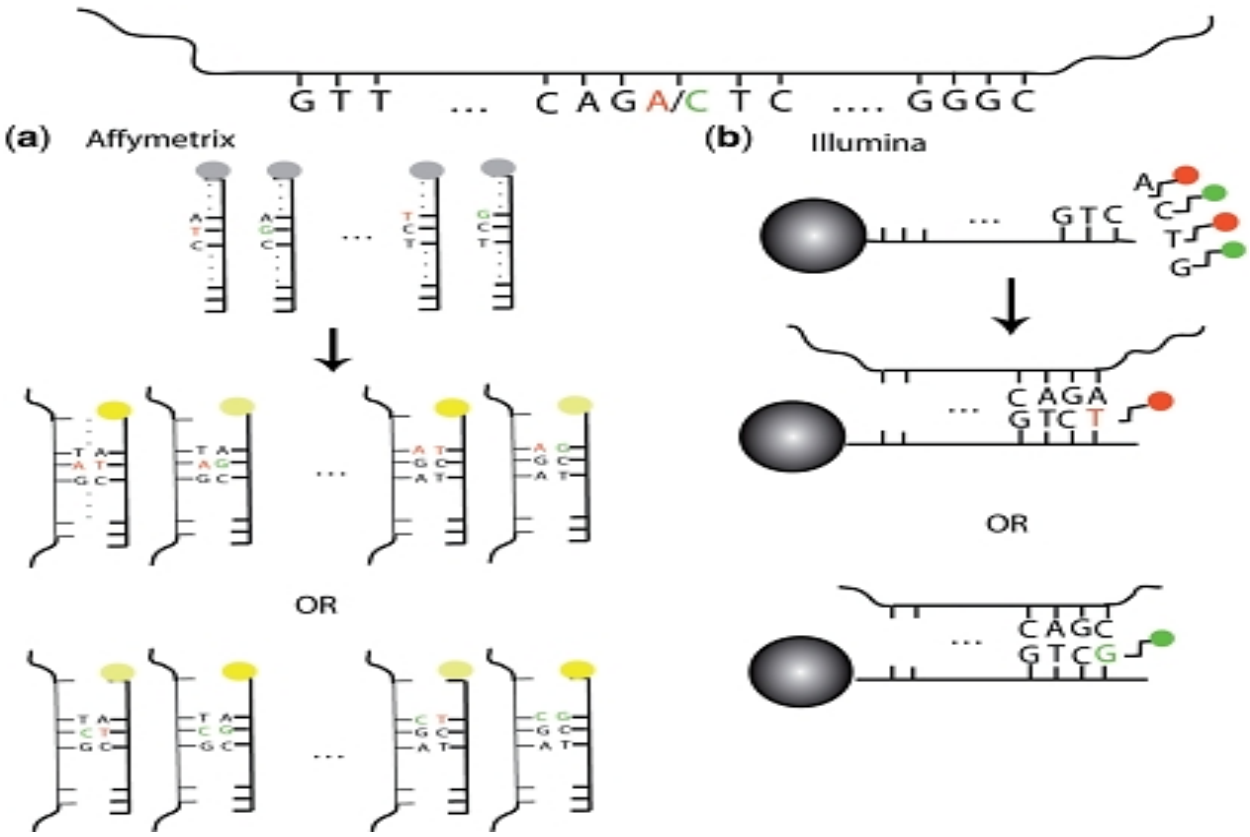

Figure 4.1: An overview of the Affymetrix (a) and Illumina (b) SNP interrogation methods. (a) Affymetrix probes are 25-mers for the interrogation of both SNP alleles, thereby allowing probe binding irrespective of SNP genotype. (b) Illumina probes are 50-mers of sites adjacent to the SNP of interest. SNP interrogation is performed through single base extension complementary to the allele included on the DNA strand. Adapted from [298].

Use of the MEGA represents a unique opportunity to assess the contribution of common SNPs to susceptibility with TBM. As mentioned previously, these SNPs often display a minor allele frequency above 0.05 with moderate to low effect sizes contributing to disease associations [166]. As the name implies common SNPs are more frequently observed in the population. This therefore requires a smaller sample size to attain a similar study power than would be required for rare SNPs. For this arm of the study, a GWAS was performed to determine if any of the SNPs present on the MEGA array are associated with susceptibility to TBM. This study represents the first GWAS performed on a TBM patient cohort.

A key concern in TBM research concerns the initial phase of pathogenesis. Crucially the method of entry for TBM has not yet been elucidated. Hypotheses for the predicted entry routes range from nasal entry bypassing the lungs and therefore being independent of pTB, to pulmonary

dissemination following pTB infection [20]. To gain further insight into possible SNP-mediated effects that may contribute to TBM development, a GWAS was performed between TBM patients and pTB patients using the Illumina® MEGA.

4.1.1 Aims

1. To perform a GWAS between TBM patients and healthy control individuals to identify SNPs associated with susceptibility to TBM.
2. To perform a GWAS between TBM patients and pTB patients to determine if any SNPs are associated with progression from active pTB to TBM, and therefore possibly aid in dissemination to the CNS.

4.2 Methods and Materials

4.2.1 Sample selection for genotyping

A total of 1 000 DNA samples were genotyped on the Illumina MEGA platform (Illumina San Diego, CA, USA). These were divided into three groupings based on disease status: the entire cohort of 123 TBM samples, 400 pulmonary TB patient samples and 477 healthy controls (Table 2.1). The 10 cases and controls used for exome sequencing were included in the MEGA genotyping sample pool. With the available sample size, we had 82% power to detect an association with an odds ratio of 5, a MAF of 0.01, and a disease prevalence of 1% as per the CaTS power calculator [291]. PTB and healthy control samples were selected according to DNA quality metrics (see section 3.2.1). All TBM samples were included for genotyping to maximise numbers. DNA samples were diluted (60ng/µl) and sent for genotyping at the Hussman Institute for Human Genomics (HIHG) (University of Miami, Florida, USA).

Quality control measures were performed on all samples at HIHG. Samples were subjected to agarose gel electrophoresis on a 0.8% agarose gel (Conditions: 100V for 1 hour, λHind III ladder) to assess the state of DNA degradation after shipping. Any samples that displayed significant degradation that would render results questionable were replaced with fresh aliquots (17 samples). Post quality control procedures, 995 samples and 11 HIHG controls were submitted for genotyping.

4.2.2 Data Quality Control

SNP genotyping microarrays are a powerful technology enabling the simultaneous genotyping of millions of SNP positions in the human genome. However, the complicated nature of the process introduces the potential for variation that can affect data analysis and results. In order to maintain consistency during analysis, it is crucial to control for variations between datasets using quality control procedures to ensure reproducible and high-quality data [292].

Genotype clusters were called on a per SNP basis at HIHG using cluster files generated by the Population Architecture using Genomics and Epidemiology (PAGE) II consortium in the genotyping module of Genome Studio version 2011.1 (Illumina, San Diego, CA, USA). Additionally, filters for SNPs with a call rate less than 85%, more than 2 replicate errors, cluster

separation values less than 2% or heterozygosity rates greater than 80%, were excluded. The genotypes were called by collaborators at Stanford University (CA, USA) and genotyping data for 1 552 576 SNPs was received.

During genotypic clustering analysis, 10 individuals were identified as having poor quality genotyping (GenCall score > 0.15) and were subsequently removed from the analysis. Using functions of PLINK v 1.07 [200], internal genotyping controls and pTB cases were removed from the dataset and new files created containing the TBM cases and healthy controls. This process left a total of 118 TBM cases and 440 healthy control individuals with genotyping information for quality control procedures. To remove single instances of duplicated SNPs the .bim file was sorted and unique values extracted. Improperly mapped SNPs are recorded as mapping to chromosome 0 by default and were removed using PLINK v1.07 [200].

Sex discordance between the user-defined .fam file and the genotyping information can be a confounding factor in downstream analyses, as males are more susceptible to TB [293, 294]. Sex discordance was calculated on a per-individual basis using PLINK v1.07 [200], where individuals with discordant sex information were corrected based on sex chromosome genotypic data [292]. Individuals with persistent sex discordance between the datasets were removed (n = 12). A search for duplicated samples was conducted with none found. As only autosomal regions are of interest in this study, sex chromosomes were removed using PLINK v1.07 [200], 49 865 SNPs were removed. After pre-QC dataset analysis, 1 465 892 SNPs in 116 TBM cases and 430 controls remained, of which 190 were male and 356 were female.

4.2.2.1 Sample QC

QC procedures for data on a sample level are critical to ensure consistency and quality of data for every individual in the study. Several quality control measures are therefore assessed to prevent the inclusion of outliers and thereby introduce sources of bias to the study. Issues such as insufficient sample concentration, contamination or labelling errors are all possible confounders in a genome-wide study and thus need to be controlled for during sample QC.

Missing information can lead to the introduction of outliers and the relative levels of missingness serve as a good indicator of overall genotyping quality [292]. Missingness assessments were carried out in PLINK v1.07 [200] at both a per-sample and per-SNP basis (see 4.2.2.2 SNP QC).

A maximum level cut-off of 0.1, corresponding to 10% as a proportion of missing information (90% genotyping rate) was used as a filter – no individuals were removed.

Heterozygosity is a relative measure to compare heterozygous vs homozygous loci per individual. Inflated measures are often indicators of possible sample contamination. Conversely, reduced levels of heterozygosity are indicative of possible consanguinity [295]. Cut-off levels of 2 standard deviations from the mean are recommended. However, due to the tight clustering of heterozygosity levels, 0.17 and 0.225 were used as more stringent minimum and maximum cut-offs, respectively. No individuals were removed from the analysis using these levels.

Related individuals introduce bias through overrepresentation of certain genotypes and will therefore skew both the allele frequencies in the sample population and the genotypic distributions observed. Identification of related individuals is determined by the average proportion of shared alleles between individuals using identity by state (IBS) calculations. The dataset was pruned for linkage disequilibrium (LD), a phenomenon where SNPs do not segregate independently during recombination, in a 50 kilobase (kb) window at a maximum r^2 correlation of 0.2 prior to IBS calculation [296]. The expectation is that a greater proportion of alleles that are identical by state will be shared between related individuals, implying identity by descent (IBD). Monozygotic twins and duplicates in the dataset would show an IBD of 1, first-degree relatives would show an IBD of 0.5 and IBD for second-degree relatives would be 0.25 [292]. Only 1 individual from a pair of related samples that showed IBD above 0.185 was removed. This IBD threshold represents a boundary halfway between second- and third-degree relatives. Using PLINK v1.07 [200], 37 individuals were removed, leaving 114 cases and 395 controls for downstream analysis.

Population homogeneity is critical when conducting GWAS as differences in ancestral population contributions between individuals may skew association results due to differing allele frequencies [292]. Assessments of population homogeneity are carried out using principal component analysis (PCA). The aim of PCA is to produce uncorrelated principal components from a dataset of variables that may be correlated with one another, with the first principal component accounting for the greatest proportion of observed variation in the dataset [297]. Eigenvectors were calculated from the data using PLINK v1.9 and the first 4 principal components plotted against one another [298]. Individuals of similar ancestral contribution produced clusters on the resultant PCA plot and cluster definition between cases and controls were compared. Slightly increased variation was

observed in the case dataset compared to controls as a result of the differences in ancestry between case individuals from different populations, however, none deviated sufficiently to be removed from downstream analyses. Thus 114 cases and 395 controls were retained in the analysis of whom 177 were male and 332 female.

4.2.2.2 SNP QC

Per-SNP QC ensures the use of only highly accurate genotyping data, consistent across all study subjects. To this end, sufficient study power is required to identify and investigate potential SNP associations. Study power is dependent upon the sample size used in the experiment and directly influences the MAF thresholds that are able to be investigated using the given sample size [299]. Therefore an increase in sample population increases study power and allows for the investigation of SNPs with lower MAFs [300]. MAF calculations were performed using PLINK v1.07 [200] across all SNPs and a filter to retain all SNPs above 5% MAF was used. This procedure removed 769 140 SNPs from the analysis, retaining 696 752 SNPs. This is expected, due to the number of low frequency SNPs found on the MEGA.

To retain a high genotyping rate and therefore confidence in genotypic calls, SNPs with missing genotype data were removed from the analysis. This reduces the potential for the discovery of false positives during association testing [292]. SNP missingness was previously calculated during sample QC using PLINK v1.07 [200] and a threshold of 0.1 missingness was applied to retain only SNPs with higher than 90% genotyping call rate. This removed 9 659 SNPs to retain 687 093 SNPs in the dataset.

In study settings where case and control study samples have been drawn from different areas, it is common practice to assess missingness differentially between the two groups as significant deviations may introduce bias [292]. Cases and controls were split and their relative SNP missingness assessed using PLINK v1.07 [200] at a threshold of 0.1, as previously discussed. Differentially missing SNPs (n = 6 597) were removed from both datasets using PLINK v1.07 [200] from both datasets to reduce the likelihood of confounding association results, leaving 680 496 SNPs.

Hardy-Weinberg Equilibrium (HWE) on a per-SNP basis was considered as severe deviations from its fundamental principles can point towards genotype calling error [292]. However, certain SNPs may be under the effects of selective pressure in the case group and therefore may contribute

to susceptibility. Therefore only the control group was assessed for severe HWE deviations using a significance threshold of $p < 0.05$ [292]. HWE calculations were performed using PLINK v1.07 [200]. This omitted 54 790 SNPs that deviated significantly, leaving 625706 SNPs for analysis at a total genotyping rate of 0.998 in 114 cases and 395 controls of which, 177 were male and 332 female.

4.2.3 Admixture Analysis

The inheritance of particular genomic portions differs between individuals as a function of the contributions to their pedigree by ancestral populations [301]. In admixed populations, genetic contributions from ancestral populations vary between sampled individuals affecting allele frequencies. This can produce false-positive or false-negative association signals if not accounted for in statistical testing [302]. Admixture mapping was performed using Admixture v1.3.0 [303] and the proportions calculated included as covariates during association testing.

The samples selected in this study were selected from different populations with varying ancestral population contributions, warranting statistical correction. Using self-reported population information and mother-tongue language information, proxy populations for the sample set were selected. Proxy populations were required due to the lack of an appropriate SAC reference dataset. These populations were selected based upon previous population genetic analyses of the SAC to determine the best ancestral populations, thus principal component analysis was not used for correction of downstream analyses [154, 304, 305]. This approach has been used in previous GWAS studies using this population [306]. PCA analysis was, however, conducted by merging study individuals with reference samples from all the available populations of the 1000 Genomes Project (Appendix 3) [307]. Downstream analyses were corrected for using ancestry proportions calculated using Admixture v1.3.0 [303]. Due to the SAC being the predominant population in both the case and control groups, five proxy populations were chosen for use in admixture calculations to account for the 5-way admixture in the SAC. The five proxies were: Khomani San from Upington, South Africa (Collaboration with Dr. Henn, Stony Brook University) Caucasian Europeans from Utah, USA (CEU), Luhya from Webuye, Kenya (LWK), Gujarati Indians from Houston, Texas, USA (GIH) and Han Chinese from Beijing, China (CHB) from HapMap phase 3 data [308]. The five proxy populations were also useful for case-group individuals who self-

identified as either of Caucasian or Xhosa origin as the CEU and LWK could be used as proxies, respectively.

Proxy individuals were merged with study individuals using PLINK v1.07 with each proxy population being represented by equal numbers of individuals [200]. SNP overlap was determined between the proxy and study individual datasets using the R programming environment [201] with only the overlapping 198 689 SNPs retained. Strand flips were identified between SNPs of the two datasets. These were rectified using the flip scan function of PLINK v1.07 [200] and the datasets were merged. Fifteen SNPs were removed since these strand flips could not be corrected due to the presence of three alleles. Genotyping quality was highlighted as being 0.54 due to the amount of missing data introduced when both datasets were combined. This was due to the use of different genotyping platforms used between the proxy and study datasets. To improve the genotyping quality, MAF filtration above 5% and missingness assessments (missingness rate >0.1) were performed using PLINK v1.07 [200] improving quality to 0.99. To increase computational efficiency during admixture calculations the assessment of independent SNPs, LD pruning using a 50kb window, a 10bp step-wise shift and an r^2 cut-off of 0.2 was performed.

Admixture v1.3.0 was run using a supervised analysis wherein reference HapMap individuals were assigned to population groups prior to analysis in an additional .pop file [303]. This allowed for the use of reference samples as a training dataset for the resultant analysis and enables efficient admixture calculation of study individuals. A default cross validation measure of 5 was used and a K-value of 5 was used to denote the assumed 5 ancestral populations of the study samples. Ancestral proportion estimations were later used as covariates in association analysis to account for the differing ancestral proportions between study individuals (Figure 4.2).

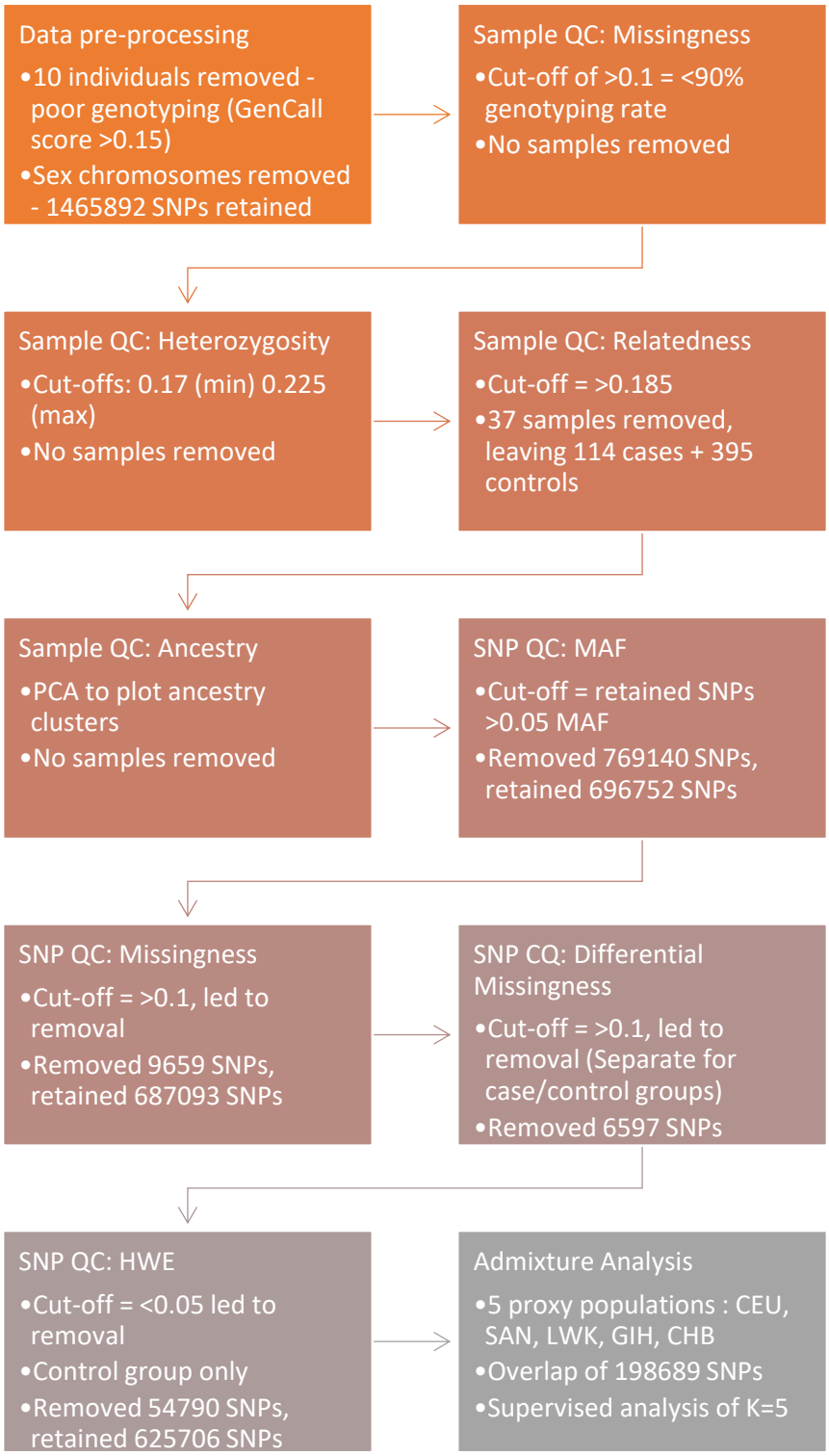

Figure 4.2: Overview of quality control procedures used during MEGA SNP genotyping array SNP prioritisation. Quality control cut-offs and procedures are included for both sample and SNP QC as well as admixture analysis.

4.2.4 Genome-Wide Association Analysis: TBM cases vs. Healthy Controls

The basis for determining statistical association of a SNP with a disease hinges upon the differences in allele frequency between case and control groups [190]. This does not imply causality, however, as the SNP of interest may be in LD with the truly causal SNP. Sample heterogeneity has to be controlled for prior to association analysis to prevent confounding of association results [309].

To that end, a covariates file containing information pertaining to potential confounding factors such as gender and ancestral proportions per-sample was created for use during analysis. In the case of ancestral proportions, only the 4 ancestries with the largest contributions were included as covariates, as including all 5 would induce perfect separation of the data. The East Asian ancestry (GIH) was excluded, as its average of 4.6% across all study individuals was the lowest. Age was not included as a covariate in the analysis due to the separation it induces in the data as cases were significantly younger than controls.

To incorporate all covariates into the association testing model, a logistic regression procedure was used to test for association under the additive model (see Figure 4.3). Logistic regression models were created with the log odds of disease representing the response variable which was predicted by additive combinations of genotype values and covariates [310]. Logistic regression was performed using PLINK v1.07 incorporating the 4 contributing ancestral populations and gender as covariates in an additive model [200]. Additive models assume a disease model where additional copies of the alternate allele increases the log odds ratio by an equal, additive amount [311].

During association testing many hypotheses are tested simultaneously, increasing the likelihood of false-positives (Type-I error). Therefore, corrections for multiple testing should be done. Multiple testing adjustment was done using PLINK v1.07 [200] during association testing. The correction procedures used independently were Bonferroni adjustment, Sidak single-step and step-down procedures, Holm-Bonferroni step-down procedure and both Benjamini and Hochberg and Benjamini and Yekutieli False Discovery Rate (FDR) measures. To account for the number of tests performed, a significance level cut-off of 7.99×10^{-8} was used.

Following the analysis, the top hits were investigated to glean an understanding of the genes and regulatory regions that may be affected by changes to the nucleotide sequence. This was done

using VEP accessible via the Ensembl genome browser [216]. Additionally, transcription factor binding site annotations were accessed using the UCSC Table Browser facility (Figure 4.3) [312].

4.2.5 Genome-Wide Association Analysis: pTB cases vs TBM cases

One of the key unanswered questions in the field of TBM research hinges on the precise method whereby *M.tb* is contained in the lung of patients who develop pTB while in a subset of active TB cases, dissemination to the bloodstream and eventually the CNS occurs. The host genome is likely to influence this phenomenon. A dataset containing the 114 post-QC TBM cases was compared to the 382 post-QC pTB cases both genotyped on the Illumina® MEGA (San Diego, CA, USA). Both datasets were subjected to identical QC procedures as outlined previously (section 4.2.2 Data Quality Control). Differential missingness calculations were performed to improve the genotyping quality to above 90% in both datasets individually, leading to the removal of 251 SNPs from the pTB dataset and 0 from the TBM dataset. PTB cases were then included in the analysis as 'controls' to facilitate association using PLINK v1.07 [200]. After merging using PLINK v1.07 [200], 271 males and 225 females were present in the dataset, with genotyping information for 575 742 SNPs. Covariates for gender and the 4 ancestral proportions were included as before.

GWAS was subsequently performed upon the merged dataset under an additive model as implemented in PLINK v1.07 (Figure 4.3) [200]. The analysis was adjusted for gender and the 4 ancestry proxies with the largest proportions: European (CEU), San, Luhya from Webuye, Kenya (LWK) and Gujarati Indians from Houston, Texas, USA (GIH). Age was not adjusted for as a covariate for reasons outlined previously. Odds ratios were calculated along with their corresponding 95% confidence intervals on a per-SNP basis. All results were adjusted for multiple testing as before during analysis. A significance cut-off of $p = 8.68 \times 10^{-8}$ was used.

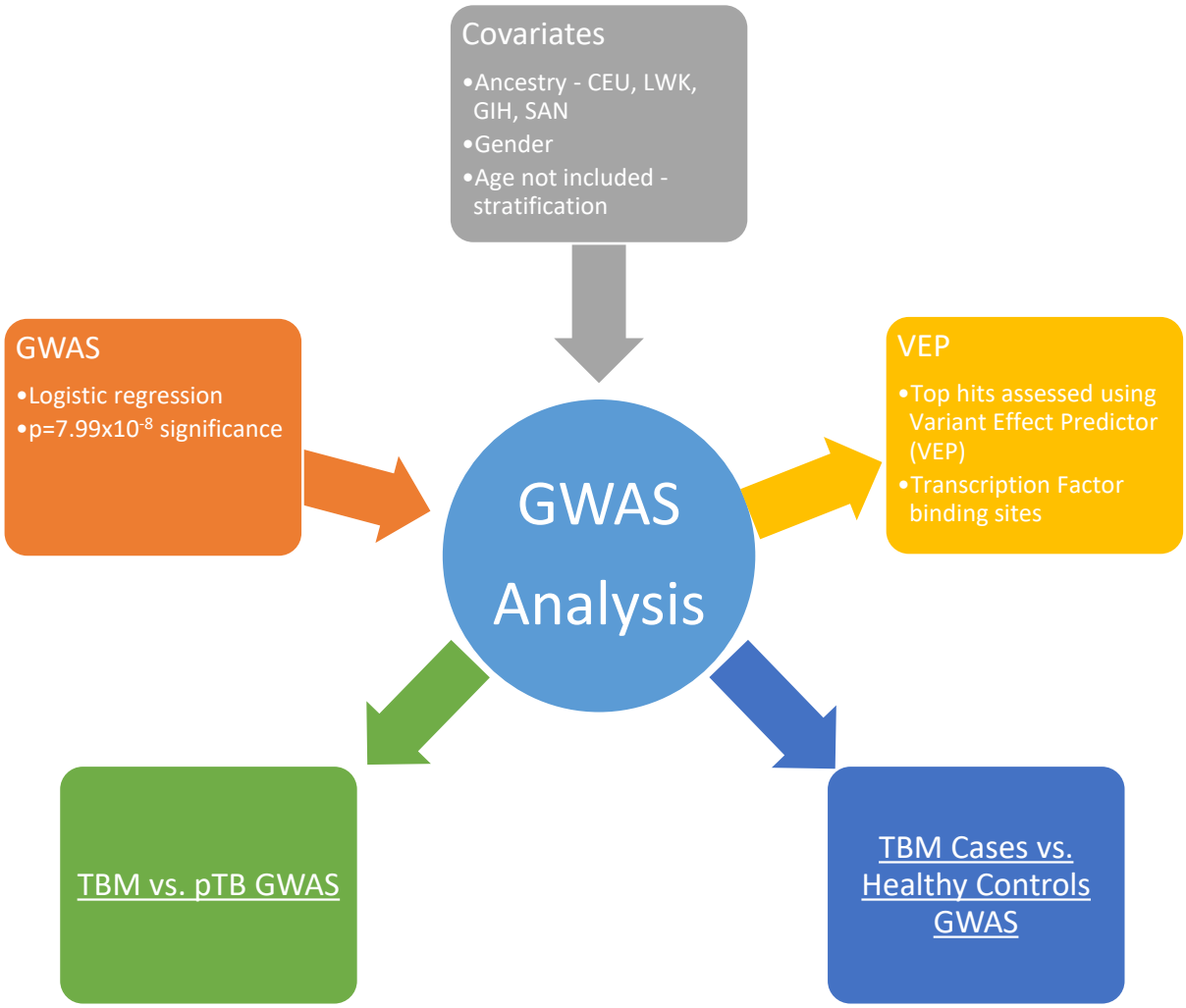

Figure 4.3: Procedures undertaken to set up the logistic regression models for the 2 GWAS analyses. GWAS = Genome-Wide Association Study, VEP = Variant Effect Predictor, CEU = Europeans from Utah, USA, LWK = Luhya from Webuye, Kenya, GIH = Gujarati Indians from Houston, Texas, USA, SAN = San individuals from Upington, South Africa

4.3 Results

4.3.1 GWAS: TBM cases vs. healthy controls

No significant associations with TBM susceptibility were found. The lowest p-value was that of rs77857429 ($p = 5.68x10^{-6}$ [OR = 4.28; 95% CI: 2.28-8.02]), an intronic SNP located in the Glucosylceramidase Beta 3 (*GBA3*) gene. Given the hypothesis generating nature of the study, an exploration of the top hits of the data was conducted. This was done to minimise false negative associations, but does expose the results to false positives. Surrounding genes and possible regulatory regions at each SNP were investigated to determine if any SNP localised to a regulatory region may affect the binding efficacy of any transcription factors, thereby influencing gene expression. The 10 SNPs with the smallest unadjusted p-values are shown in Table 4.1.

Table 4.1: Exploratory dataset for the TBM cases vs. healthy controls GWAS

CHR[a]	SNP	Base-pair Position	Unadjusted p-value	Genomic Control	Bonferroni Adjusted p-value	Odds Ratio [95% CI][b]
4	rs77857429	22804085	$5.68x10^{-6}$	$7.17x10^{-6}$	1	4.28 [2.28-8.02]
17	rs3760495	684729	$9.17x10^{-6}$	$1.15x10^{-5}$	1	2.72 [1.75-4.24]
17	rs2273454	685640	$9.17x10^{-6}$	$1.15x10^{-5}$	1	2.72 [1.75-4.24]
17	rs2750007	677226	$1.50x10^{-5}$	$1.85x10^{-5}$	1	3.94 [2.12-7.33]
3	rs17046322	6952739	$1.82x10^{-5}$	$2.24x10^{-5}$	1	2.7 [1.71-4.25]
4	rs1511144	127535445	$2.03x10^{-5}$	$2.49x10^{-5}$	1	2.27 [1.56-3.3]
10	rs114255276	14825449	$2.03x10^{-5}$	$2.5x10^{-5}$	1	2.97 [1.8-4.89]
18	rs6565824	73487088	$2.25x10^{-5}$	$2.75x10^{-5}$	1	0.41 [0.27-0.62]
22	rs1127000	50716167	$2.44x10^{-5}$	$2.99x10^{-5}$	1	2.24 [1.54-3.26]
12	rs10774604	111168743	$2.57x10^{-5}$	$3.14x10^{-5}$	1	3.67 [2-6.73]

a. CHR = Chromosome
b. CI = Confidence Interval

Four of the SNPs listed in the top hits were localised in transcription factor binding sites. Table 4.2 shows the gene each SNP is localised to including the type of SNP and any transcription factor binding sites that it may affect.

SNP	Gene Symbol	Type of SNP	Transcription factor binding sites
rs77857429	*GBA3*	Intronic	N/A
rs3760495	*GLOD4*	Intronic	*CTCF, RAD21, SMC3*
rs2273454	*GLOD4*	Upstream of gene	*POLR2A, SAP30, TCF3*
rs2750007	*GLOD4*	Intronic	*FOXM1, RELA*
rs17046322	*GRM7*	Intronic	N/A
rs1511144	*LOC102724210/RBM48P1*	Intergenic	N/A
rs114255276	*FAM107B*	Promoter Flanking	*GTF3C2*
rs6565824	*LOC100505853*	Promoter Flanking	N/A
rs1127000	*PLXNB2*	Splice site	N/A
rs10774604	*PPP1CC*	Intronic	N/A

Table 4.2: Localisation of each of the top hit SNPs from the GWAS: TBM cases vs. healthy controls

None of the highlighted SNPs were found to be located in exons and all were either intronic or intergenic. The Plexin B2 gene (*PLXNB2*) encodes several transcripts and, depending on the transcript, the rs1127000 SNP is either a splice site variant or a synonymous SNP. However, using Variant Effect Predictor (VEP) software, rs1127000 was predicted to be of low effect when all transcripts were considered [216].

4.3.2 GWAS: pTB cases vs. TBM cases

No significant associations were identified when comparing the two cases groups to assess progression from pTB to TBM. The SNPs listed in Table 4.3 are the top hits of the analysis and using a significance threshold of $p=8.68x10^{-8}$ no SNPs are close to the significance threshold. The SNP with the lowest p-value was rs2679308, located in an intron in *LOC102724084*, an uncharacterised gene. The p-value of rs2679308 was $p=5.11\ x10^{-6}$.

A dataset of the top hits of the analysis was then used for exploratory purposes. This increased threshold means that the likelihood of false positives is high. This dataset in Table 4.3 serves as a safe-guard to avoid false negative results and may provide additional SNPs for genotyping in larger cohorts.

Table 4.3: Exploratory dataset for the TBM cases vs. pTB cases GWAS.

Chr[a]	Base-pair Position	SNP	Unadjusted p-value	Genomic Control	Bonferroni Adjusted p-value	Odds Ratio [95% CI][b]
16	80350711	rs2679308	$5.11x10^{-6}$	$7.57x10^{-6}$	1	26.43 [6.47-108]
17	51638548	rs4309447	$5.72x10^{-6}$	$8.44x10^{-6}$	1	3.70 [2.10-6.50]
2	121302719	rs2140779	$5.95x10^{-6}$	$8.78x10^{-6}$	1	18.12 [5.17-63.5]
17	51653566	rs2055478	$8.29x10^{-6}$	$1.21x10^{-5}$	1	3.61 [2.05-6.34]
5	1442842	rs13189021	$9.74x10^{-6}$	$1.41x10^{-5}$	1	11.66 [3.93-34.61]
7	138601826	rs2251220	$1.31x10^{-5}$	$1.89x10^{-5}$	1	3.66 [2.04-6.57]
13	112048353	rs61966912	$1.85x10^{-5}$	$2.61x10^{-5}$	1	19.4 [4.99-75.35]
4	166976065	rs1393854	$1.93x10^{-5}$	$2.72x10^{-5}$	1	8.61 [3.21-23.11]
4	168490955	rs1914621	$1.97x10^{-5}$	$2.79x10^{-5}$	1	0.33 [0.20-0.55]
1	168987665	rs61803268	$2.02x10^{-5}$	$2.85x10^{-5}$	1	8.45 [3.17-22.54]

a. CHR = Chromosome
b. CI = Confidence Interval

Table 4.4 shows the exploratory dataset when comparing pTB and TBM cases with gene, SNP type and transcription factor binding site annotations. A single exonic SNP, rs2251220, was in the exploratory dataset and was localised to *KIAA1549*. The rs2251220 SNP induced a non-synonymous change from a Serine to Leucine in the resultant protein. VEP analysis of rs2251220 using SiFT and Polyphen plugins showed that the SNP was tolerated and benign with scores of 0.1 and 0.053 respectively. The rs4309447 and rs2140779 SNPs were located in transcription factor binding sites. The rs2140779 SNP was located in a transcription factor binding site upstream of the Glioma-Associated Oncogene Family Zinc Finger 2 (*GLI2*) gene and is bound by the Jun transcription factor. This suggests a possible role in both the cis- and trans-regulation of several genes.

Table 4.4: Localisation of each of the top-hit SNPs from the GWAS of TBM cases vs. pTB cases

SNP	Gene	SNP Type	Transcription Factor Binding Sites
rs2679308	*LOC102724084*	Intronic	N/A
rs4309447	*LOC100419014/RPS2P48*	Intergenic	*FOS, JUND*
rs2140779	*LOC84931/GLI2*	Upstream SNP	*JUND*
rs2055478	*LOC100419014/RPS2P48*	Intergenic	N/A
rs13189021	*SLC6A3*	Intronic	N/A
rs2251220	*KIAA1549*	Exonic	N/A
rs61966912	*TEX29/LOC101928616*	Intergenic	N/A
rs1393854	*TLL1*	Intronic	N/A
rs1914621	*SPOCK3/PHBP14*	Intergenic	N/A
rs61803268	*LINC00970*	Intronic	N/A

4.4 Discussion

4.4.1 GWAS – TBM vs. Healthy Controls

The advent of GWAS has enabled researchers to highlight regions of the genome that confer susceptibility towards certain diseases. These SNPs may not be causative for the disease as they may tag the causative SNP. However, this procedure greatly focusses the search area for truly causative SNPs in the human genome. As yet, GWASs have not been undertaken in a TBM cohort and therefore susceptibility loci linked specifically to TBM are currently few. Many of these are known as a result of candidate genes that have been extrapolated from pTB. This approach is flawed as the two diseases follow different mechanisms of pathogenesis, despite sharing the same infectious bacterium.

To gain a better understanding of susceptibility loci specific to TBM we performed a GWAS using the Illumina MEGA. Using logistic regression for SNP testing and correction for multiple testing and covariates, no significant genome-wide associations with TBM disease were detected. Examination of the top-hits in the analysis showed that 3 of the top 4 SNPs were located in or around the Glyoxalase Domain Containing 4 (*GLOD4*) gene. In addition to this, all 3 displayed odds ratios pointing towards being susceptibility factors. As yet, *GLOD4* has no known function and has not been investigated in the context of TBM. Further investigation into their biological relevance showed that rs3760495 and rs2750007 were intronic and rs2273454 was located upstream of *GLOD4* (Table 4.2). All 3 of these SNPs were located in transcription factor binding

sites bound by multiple transcription factors. Of particular interest was rs2273454 which was found to be located in a binding site for RNA polymerase II. This points to the SNP being located in the *GLOD4* promoter region and may affect gene expression. As a result, future investigation into possible diminished RNA polymerase II binding and therefore diminished transcription of *GLOD4* as a result of rs2273454 should be investigated.

Considering the sample size used in the GWAS, this lack of association is likely attributable to the diminished study power observed when using a study population of 123 TBM patients and 477 healthy controls. This does not imply that these SNPs may not be associated with TBM in other population groups as a result of allele frequency differences or when examined in a larger sample size with greater statistical power to detect associations. This is however one of the largest TBM collections in the world and therefore represents a valuable resource for TBM research, especially considering the challenges involved in patient recruitment of such a rare disease.

4.4.2 GWAS – TBM cases vs. pTB cases

The mechanism whereby TBM disease is developed is one of the central questions of TBM research. In a subset of individuals, *M.tb* is not properly contained in the lung and disseminates to the CNS. One possible contributing factor to this phenomenon is the host genome and in particular differences between the genomes of pTB and TBM patients. Other factors such as *M.tb* strain differences and environmental factors will contribute to TBM development. *M.tb* strain differences have a proven role in their disseminatory capacity. The Beijing strain is considered to be more virulent and is the causative strain for the majority of TBM cases [109, 116]. The varying dissemination capacity between strains exposes a possible bias in this result as culture of *M.tb* from some TBM patient samples is not possible [313]. A link has also been shown between certain populations and susceptibility to certain *M.tb* strains thus further complicating the problem [314]. Natural variation of the host genome is expected between all humans, however, SNPs in key immunoregulatory genes may assist in answering this question central to TBM research.

The GWAS of TBM cases compared to pTB cases did not yield any SNPs that were associated with progression from pTB to TBM (Table 4.3). Exploratory analysis into the top hits of the GWAS revealed a single exonic SNP, rs2251220. Assessment of the substitution induced by rs2251220 using the prediction tools SiFT and Polyphen-2 determined that the non-synonymous

change from a Serine to Leucine was benign and tolerated respectively [265, 266]. SNP selection for the MEGA array was based primarily around the selection of tag-SNPs therefore the location of the SNP being investigated is often not the true location of the SNP that is driving the effects seen in this study. Therefore one cannot discount that these SNPs may be associated with TBM progression in other populations with different allele frequency spectrums. In that case, the LD structure surrounding each SNP would require investigation to determine the SNP truly driving the association and determine its biological relevance.

In this study both GWAS studies suffered from low sample numbers and were particularly affected by the age difference between the cases and controls. This may have impacted upon the results of the study and may be a contributing factor to the negative results of the logistic regression. In addition, the use of the full TBM cohort for the GWAS introduced some fine-scale population structure as the cases were not all from the SAC and some were from the Xhosa population. Previously covariates that were correlated with the phenotype of interest were corrected for by including these in logistic regression models [315]. Logistic regression testing using case and control cohorts are vulnerable to confounding by population structure and stratification, the result being an increase in test statistics and a higher false positive rate [316]. However, in studies using binary or categorical variables where the selection of cases and controls leads to a correlation between the covariate and the genotype of interest, a reduction in study power will ensue due to an increase in the standard error [315]. Controlling for population structure in logistic regression is traditionally done through the inclusion of principal components as covariates during model construction, this considered as a broad scale adjustment [317]. Likewise controlling for the genomic inflation factor, a measure of inflation of the test statistic, aids in reducing the false positive rate. Adjustment for the genomic inflation factor is recommended when a homogenous population is used in the study and no significant differentiation is seen [317]. In the case of a stratified sample set like the one used here, a mixed model approach should be considered.

Mixed linear models such as the variance component approach, efficient mixed-model association (EMMA), are often considered to be computationally intensive in datasets consisting of large sample numbers. This inefficiency is as a result of the need for estimation of variance parameters [318]. Optimisation of this process was achieved through the estimation of variance parameters only once per dataset as often the effect of a particular locus is often upon a trait is marginal [319].

This optimised approach is known as EMMA eXpedited (EMMAX) and is able to account for population structure through measures of pairwise relatedness between individuals [317]. The structure is controlled for through the use of an empirically calculated relatedness matrix as a correlation of phenotypes is expected in genetically similar individuals [318]. This approach has been previously applied to whole genome association studies [320–322]. Through use of a mixed model approach such as EMMAX, the study power would be increased and thorough correction for all covariates and age stratification would be possible. This represents a viable solution for the structure involved in studies involving an early onset rare disease which is considered a challenge for patient recruitment.

Concluding Remarks Chapter

5

Chapter 5: Concluding remarks

This study contributes significantly to the TB host genetics field, since we present the first exome sequencing study and GWAS of children with a rare form of TB, namely TBM. The overarching aim of this study was to assess the contribution, if any, of common SNPs of varying frequency to TBM susceptibility. To this end, SKAT gene-set based analyses were used to assess the contribution of rare SNPs (MAF < 0.01) and their cumulative effect when combined with common SNPs (MAF > 0.01) from exome sequencing data. Two GWAS were performed on the Illumina® MEGA including samples from TBM cases, pTB cases and healthy controls. It is also the first exome sequencing and GWAS study to be performed on TBM patients from South Africa, giving insight into manifestations of TBM patients in this setting. Exome sequencing of the healthy controls will provide a useful resource for other researchers working with the SAC population.

Our GWAS studies were designed to identify susceptibility loci for TBM development as either a primary infection or as a result of dissemination from a primary pulmonary focus. Neither of these analyses, however, identified any loci associated with TBM in the SAC. This is likely to be a direct result of insufficient study power due to small sample size.

An association with TBM susceptibility was with *CCP110* ($p = 5.89x10^{-6}$) (rs3751821, rs226891, rs7190666) in the SKAT Common Rare analysis which assessed the cumulative effect of common and rare variants between TBM cases and healthy controls. This association, however, is possibly a false positive due to the insufficient covariate correction, but may also point towards a previously unknown role for *CCP110*. No significant associations were found with TBM infection using the SKAT-O gene-set association test or for the GWAS between TBM cases and healthy controls.

The combination of all these analyses provided a number of candidate genes for future study in the context of TBM, some of which may lead to advances in our current understanding of TBM pathogenesis. These are highlighted in Table 5.1 and can in future be studied both genetically and functionally.

Table 5.1: The candidate genes identified in this study

Gene	Association Test	Unadjusted p-value	Relevance
ZFHX3	SKAT-O	4.63×10^{-4}	SKAT-O, Lowest p-value
CYP4F2	SKAT-O	1×10^{-3}	LTB_4 regulation
NOD2	SKAT-O	6.71×10^{-3}	TNF-α regulation and pathogen recognition
CCP110	**SKAT Common Rare**	**5.89×10^{-6}**	**SKAT Common Rare lowest p-value**
A2M	SKAT Common Rare	8.63×10^{-3}	Cytokine transporter
CX3CR1	SKAT Common Rare	9.57×10^{-3}	T and B cell recruitment, microglia response
MARCO	SKAT Common Rare	1.75×10^{-3}	Pathogen recognition
ROR1	SKAT Common Rare	6.60×10^{-4}	Regulation of tight junctions
ATG10	SKAT Common Rare	8.56×10^{-4}	Autophagy – phagosomal maturation
IFITM2	SKAT Common Rare	9.23×10^{-3}	Innate immunity – signal transduction
C4BPA	SKAT Common Rare	2.54×10^{-4}	Complement system regulation
IRAK-M	SKAT Common Rare	8.24×10^{-3}	TLR pathway regulation

Significant results are highlighted in bold text

The innate immune system consists of a series of interacting pathways which cumulatively provide the human host with protection against invading pathogens. Due to the complex interactions involved in immunity, a certain level of redundancy between pathways is expected. This redundancy acts as a compensatory measure if defects in one pathway is present. While many of the genes listed in Table 5.1 are part of the innate immune system, their functional annotations are varied and not localised to any one pathway of innate immunity. Thus if any defect in these genes were to arise it would likely be compensated for by interacting genes. It is these redundancy measures which may explain why no significant associations were found in some analyses. We therefore recommended that the total complement of genes acting in a single pathway be investigated for association with disease to mitigate the effects of redundancy. This will allow

proper assessment of the role these genes play in innate immunity and further our knowledge of TBM pathogenesis.

This study had several limitations which may affect the conclusions that can be drawn from results or perhaps hinder the detection of results in the first instance:

1. Patient recruitment for the rare TBM disease phenotype is a major challenge and this resulted in a small sample size. This means the study did not have sufficient power to detect associations with variants with lower allele frequencies (1 – 5%) and small effect sizes. This will particularly influence the SKAT-O analysis as rare SNP association studies require large sample numbers in order to be sufficiently powered to detect associations with rare SNPs of small effects [171, 299, 323, 324]. Sample size impacted on covariate correction as well. Consequently correction for ancestry and age was not possible in the two SKAT analyses as inclusion of more than 1 covariate besides gender in the logistic regression induced perfect separation in the analysis. Therefore, it is possible that our results are enriched for Type 1 errors and these results should be validated in additional and larger sample sets.
2. During quality control procedures for the MEGA SNP genotyping array, filtration for the retention of SNPs with an MAF >0.05 was performed. This was done in consideration of the sample size with the aim of retaining sufficient study power. In doing so, 499 859 SNPs were removed from the analysis. This represents a significant proportion of the 1.7 million SNPs that were genotyped originally. SNP selection methods used during the creation of the array likely influenced the number removed as the allele frequencies from studies on different populations were taken into account during SNP selection. Due to the differences in allele frequency between populations and the admixed nature of the SAC population, it is likely that the 769 140 SNPs removed were of very low frequency in the SAC population compared to populations used in SNP selection.
3. Age stratification between cases and controls during sample selection will have introduced bias into the analysis. This, however, was unavoidable as the most susceptible age for TBM onset is between 0-5 years of age. In contrast, control samples needed to be of sufficient age to decrease the chance of TBM development. Additionally, control individuals required sufficient exposure to pTB to ensure that active pTB infection did not occur in these

individuals. This stratification in age was also noted in a Russian study of pTB wherein the cases were significantly younger than the control group [325]. Age stratification is often necessary in childhood diseases as a result of disease pathogenesis. A control group that is older often displays a lower risk for the phenotype of interest allowing for accurate comparisons. The results from both GWAS analyses were also affected by age stratification.

4. The *ROR1* SNP rs7527017, found in the top hits of the SKAT Common Rare analysis was genotyped in the MEGA genotyping data. Being a larger cohort, this represented an opportunity to assess its possible association with TBM, along with 108 other SNPs in *ROR1*. However, it was removed during quality control measures due to significant deviation from HWE (p=0.01245) in the control group. HWE deviations in the control group may in fact be as a result of the admixed nature of the SAC population and not as a result of poor genotyping [326]. This means we could not analyse this SNP in the larger sample set. In retrospect, the chosen threshold of p = 0.05 was likely too stringent and considering the MAFs in the proxy populations: CEU = 0.409, CHB = 0.32, GIH = 0.296 and LWK = 0.04 the threshold for exclusion could have been lowered to $p < 0.001$ as stated in literature [292].
5. The basis of exome sequencing is to sequence only the exonic regions of the genome which represent approximately 1.2% of the total nucleotide content [159]. The library preparation does involve the targeting of the 5'- and 3'UTR regions as well, with the aim of capturing regulatory sites. However, due to the relatively small portion of the genome that is actually sequenced, a limitation of the study is the genome coverage.
6. Population structure was present in the TBM case group as the entire patient cohort was used for GWAS. This included Xhosa and SAC individuals and therefore introduced sample heterogeneity. Population structure is commonly corrected for through the inclusion of principal components as covariates in the testing model. In this case, correction using PCA was not done as ancestry components were corrected for using Admixture v1.3.0 using pre-determined proxy populations. However, the use of proxy populations is biased and does not represent an empirical assessment of the population sub-structure.

Future studies would include the verification of the association seen between *CCP110* and TBM, preferably in a larger cohort. If this result is verified, functional studies into the mechanisms of

CCP110 can be studied in the context of TBM which may identify a novel function for *CCP110*. During exome sequencing analyses 10 candidate genes were identified which function in pathways of the innate immune system. Several of these genes play a regulatory role in the pathway and may play a critical role in the regulation of the inflammatory response, a key factor in TBM pathogenesis and neurological sequelae. The candidate genes in Table 5.1 must be interpreted with caution as our knowledge of the majority of biological pathways is fragmented at best. Several genes may not be involved in TBM pathogenesis as a result of biological interactions or relationships currently unknown. Future genotyping studies to determine the association of SNPs in the genes listed above with increased TBM susceptibility will also determine if the genes are involved in TBM. These SNPs would then be followed up with functional analyses, assuming associations are found, which could measure expression levels and their effects upon pathway function in the context of TBM disease. Age stratification was a major issue in the study design and may have impacted upon the results obtained. It may be possible to mitigate the effects of the stratification by "spiking" the control group used in each analysis with a number of case individuals. The aim of this will be to reduce the separation induced through stratification. An approach such as this will lower the overall study power, but may be counteracted through the use of a large cohort. Additionally application of Firth penalised regression models would aid reducing bias caused by separation [327]. Association analyses would then take the form of trend analyses to determine association of SNPs with TBM susceptibility.

Strategies to cope with small sample sizes and population structure such as mixed models would be an alternative to logistic regression without a significant reduction in power. Recommendations for future studies would include performing association testing of this dataset using a mixed model approach such as EMMAX as explained in section 4.4.2. The use of this will provide an unbiased correction for age stratification and population sub-structure, instead of using proxy populations for the admixture present in the SAC group.

As an alternative solution to the age stratification between cases and controls, adult TBM cases with HIV can be compared to HIV positive individuals without TBM. HIV status is a known susceptibility factor for extrapulmonary TB and pTB development [328]. Thus, HIV positive controls that do not develop TBM or TB of any form must have extreme genetic resistance factors [329]. While sampling of such a specific and extreme phenotype is possibly more challenging than

the recruitment strategy employed in this study, sampling of extreme phenotypes has been shown to increase study power in sequencing studies [162–164].

Under ideal circumstances the overall study design could have been modified to allow the use of the exome sequencing dataset as a replication cohort for findings from the GWAS. However, in this case this was not possible as the 10 TBM case samples that were used for exome sequencing were also used in both GWAS analyses. A recommendation in this case would be to perform gene-set based association testing on the larger GWAS dataset to ascertain whether findings from the exome sequencing would still be present in the larger cohort used for GWAS. An advantage of this approach would be a significant reduction in the multiple testing burden and additionally would facilitate the proper correction for covariates.

TBM remains the form of TB with the highest mortality rate and therefore represents a major problem for medicine worldwide. It is clear that not enough is known about disease pathogenesis mechanisms as critical factors such as the method of dissemination to the bloodstream and eventually the CNS are still unknown. TBM represents an extreme form of TB where the contribution of host genes and strain type are more likely to determine disease presentation than environmental factors. Our GWAS results support this hypothesis, since the TBM cases and pTB patients share similar environmental and economic circumstances. We also posit that defects during CNS development may contribute to TBM development rendering critical barriers such as the BBB or BCB vulnerable to pathogenic invasion. TB is a complex disease that has known environmental, bacterial genotypic and host genotypic factors which influence its presentation and disease severity. This is also true for TBM and certain mycobacterial strains are more often found to be the causative agent. However, it is also apparent that the host genotype plays a role in the disseminatory capacity of mycobacterial strains in patients as not all patients with these strains develop TBM. This suggests that polymorphisms could induce defective protection against infection in these patients. The subset of genes that confer susceptibility to TBM may likely overlap with those found to influence susceptibility to pTB, particularly those with involvement in innate immunity. However, differences between the susceptibility gene subsets will likely be apparent due to the inclusion of genes that may render the BBB vulnerable and others that may aid in dissemination. Through the findings of this study, it seems that SNPs of higher frequency but comparatively more moderate effect sizes influence the susceptibility of TBM patients, instead of

rare SNPs of large effect. This, however, does not imply that rare SNPs do not play a role in TBM pathogenesis. Three genes involved in innate immunity were identified during the exome sequencing analyses, namely: *IRAK-M*, *C4BPA* and *IFITM2*. Due to the immunoregulatory nature of both *IRAK-M* and *C4BPA* and the containment of *M.tb* by *IFITM2*, it is likely that these genes may play a synergistic role in TBM pathogenesis. Although they do not function in the same pathway, their cumulative effect upon the innate immune system should not be discounted. The results of association studies such as the one presented here may possibly contribute towards the identification of more susceptible individuals in the population in the future of personalised medicine. Through this, different vaccination strategies and treatment durations may be instigated to tailor treatments to the needs of the patient and hopefully protect against TBM, a disease with debilitating consequences even for survivors.

References

1. World Health Organization (2014) WHO report 2014: Global Tuberculosis Control 2014 - epidemiology, strategy, financing. Geneva

2. WHO | Global tuberculosis report 2015. In: WHO. http://www.who.int/tb/publications/global_report/en/. Accessed 2 Nov 2015

3. Jepson A, Fowler A, Banya W, et al (2001) Genetic regulation of acquired immune responses to antigens of Mycobacterium tuberculosis: a study of twins in West Africa. Infect Immun 69:3989–3994.

4. Shaw MA, Collins A, Peacock CS, et al (1997) Evidence that genetic susceptibility to Mycobacterium tuberculosis in a Brazilian population is under oligogenic control: linkage study of the candidate genes NRAMP1 and TNFA. TuberLung Dis 78:35–45.

5. UNAIDS, WHO (2009) AIDS epidemic update.

6. Chamie G, Marquez C, Luetkemeyer A (2014) HIV-Associated Central Nervous System Tuberculosis. Semin Neurol 34:103–116. doi: 10.1055/s-0034-1372347

7. World Health Organization (2009) WHO report 2009: Global Tuberculosis Control 2009 - epidemiology, strategy, financing. Geneva

8. Health Systems Trust: Incidence of TB in the provinces of South Africa. (2009) http://www.hst.org.za/healthstats/16/data. In: http://www.hst.org.za/healthstats/16/data. http://www.hst.org.za/healthstats/16/data.

9. CDC - TB Control Strategic Efforts: Botswana - Global TB - TB. http://www.cdc.gov/tb/topic/globaltb/southafrica.htm. Accessed 4 Feb 2015

10. van Leeuwen LM, van der Kuip M, Youssef SA, et al (2014) Modeling tuberculous meningitis in zebrafish using Mycobacterium marinum. Dis Model Mech 7:1111–1122. doi: 10.1242/dmm.015453

11. Well GTJ van, Paes BF, Terwee CB, et al (2009) Twenty Years of Pediatric Tuberculous Meningitis: A Retrospective Cohort Study in the Western Cape of South Africa. Pediatrics 123:e1–e8. doi: 10.1542/peds.2008-1353

12. Patkar D, Narang J, Yanamandala R, et al (2012) Central Nervous System Tuberculosis: Pathophysiology and Imaging Findings. Neuroimaging Clin N Am 22:677–705. doi: 10.1016/j.nic.2012.05.006

13. Hernandez Pando R, Aguilar D, Cohen I, et al (2010) Specific bacterial genotypes of Mycobacterium tuberculosis cause extensive dissemination and brain infection in an experimental model. Tuberculosis 90:268–277. doi: 10.1016/j.tube.2010.05.002

14. Thuong NTT, Dunstan SJ, Chau TTH, et al (2008) Identification of Tuberculosis Susceptibility Genes with Human Macrophage Gene Expression Profiles. PLoS Pathog 4:e1000229. doi: 10.1371/journal.ppat.1000229

15. Berman S, Kibel MA, Fourie PB, Strebel PM (1992) Childhood tuberculosis and tuberculous meningitis: high incidence rates in the Western Cape of South Africa. TuberLung Dis 73:349–355.

16. Maree F, Hesseling AC, Schaaf HS, et al (2007) Absence of an association between Mycobacterium tuberculosis genotype and clinical features in children with tuberculous meningitis. Pediatr Infect Dis J 26:13–18. doi: 10.1097/01.inf.0000247044.05140.c7

17. Cruz AT, Starke JR (2007) Clinical manifestations of tuberculosis in children. PaediatrRespirRev 8:107–117.

18. Visser DH, Solomons RS, Ronacher K, et al (2015) Host Immune Response to Tuberculous Meningitis. Clin Infect Dis 60:177–187. doi: 10.1093/cid/ciu781

19. Brancusi F, Farrar J, Heemskerk D (2012) Tuberculous meningitis in adults: a review of a decade of developments focusing on prognostic factors for outcome. Future Microbiol 7:1101–1116. doi: 10.2217/fmb.12.86

20. Donald PR, Schaaf HS, Schoeman JF (2005) Tuberculous meningitis and miliary tuberculosis: the Rich focus revisited. J Infect 50:193–195. doi: 10.1016/j.jinf.2004.02.010

21. Thwaites G, Chau TT, Mai NT, et al (2000) Tuberculous meningitis. JNeurolNeurosurgPsychiatry 68:289–299.

22. Rich A, McCordick H (1933) The pathogenesis of tuberculous meningitis. Bull John Hopkins Hosp 52:5–37.

23. Rock RB, Olin M, Baker CA, et al (2008) Central Nervous System Tuberculosis: Pathogenesis and Clinical Aspects. Clin Microbiol Rev 21:243–261. doi: 10.1128/CMR.00042-07

24. Chatterjee S (2011) Brain tuberculomas, tubercular meningitis, and post-tubercular hydrocephalus in children. J Pediatr Neurosci 6:S96–S100. doi: 10.4103/1817-1745.85725

25. Thwaites GE, Tran TH (2005) Tuberculous meningitis: many questions, too few answers. Lancet Neurol 4:160–170. doi: 10.1016/S1474-4422(05)01013-6

26. Abbott NJ, Rönnbäck L, Hansson E (2006) Astrocyte–endothelial interactions at the blood–brain barrier. Nat Rev Neurosci 7:41–53. doi: 10.1038/nrn1824

27. Be NA, Bishai WR, Jain SK (2012) Role of Mycobacterium tuberculosis pknD in the pathogenesis of central nervous system tuberculosis. BMC Microbiol 12:7. doi: 10.1186/1471-2180-12-7

28. Orihuela CJ, Mahdavi J, Thornton J, et al (2009) Laminin receptor initiates bacterial contact with the blood brain barrier in experimental meningitis models. J Clin Invest 119:1638–1646. doi: 10.1172/JCI36759

29. Liu T-B, Kim J-C, Wang Y, et al (2013) Brain inositol is a novel stimulator for promoting Cryptococcus penetration of the blood-brain barrier. PLoS Pathog 9:e1003247. doi: 10.1371/journal.ppat.1003247

30. Bermudez LE, Sangari FJ, Kolonoski P, et al (2002) The efficiency of the translocation of Mycobacterium tuberculosis across a bilayer of epithelial and endothelial cells as a model of the alveolar wall is a consequence of transport within mononuclear phagocytes and invasion of alveolar epithelial cells. Infect Immun 70:140–146.

31. Macgregor AR, Green CA (1937) Tuberculosis of the central nervous system, with special reference to tuberculous meningitis. J Pathol Bacteriol 45:613–645. doi: 10.1002/path.1700450312

32. Wu HS, Kolonoski P, Chang YY, Bermudez LE (2000) Invasion of the brain and chronic central nervous system infection after systemic Mycobacterium avium complex infection in mice. Infect Immun 68:2979–2984.

33. Curto M, Reali C, Palmieri G, et al (2004) Inhibition of cytokines expression in human microglia infected by virulent and non-virulent mycobacteria. Neurochem Int 44:381–392. doi: 10.1016/j.neuint.2003.08.012

34. Menozzi FD, Reddy VM, Cayet D, et al (2006) Mycobacterium tuberculosis heparin-binding haemagglutinin adhesin (HBHA) triggers receptor-mediated transcytosis without altering the integrity of tight junctions. Microbes Infect 8:1–9. doi: 10.1016/j.micinf.2005.03.023

35. Mueller-Ortiz SL, Sepulveda E, Olsen MR, et al (2002) Decreased infectivity despite unaltered C3 binding by a DeltahbhA mutant of Mycobacterium tuberculosis. Infect Immun 70:6751–6760.

36. Pethe K, Bifani P, Drobecq H, et al (2002) Mycobacterial heparin-binding hemagglutinin and laminin-binding protein share antigenic methyllysines that confer resistance to proteolysis. Proc Natl Acad Sci U S A 99:10759–10764. doi: 10.1073/pnas.162246899

37. Yorsangsukkamol J, Chaiprasert A, Palaga T, et al (2011) Apoptosis, production of MMP9, VEGF, TNF-alpha and intracellular growth of M. tuberculosis for different genotypes and different pks5/1 genes. Asian Pac J Allergy Immunol Launched Allergy Immunol Soc Thail 29:240–251.

38. Matsuyama W, Hashiguchi T, Umehara F, et al (2001) Expression of vascular endothelial growth factor in tuberculous meningitis. J Neurol Sci 186:75–79.

39. Jain SK, Paul-Satyaseela M, Lamichhane G, et al (2006) Mycobacterium tuberculosis Invasion and Traversal across an In Vitro Human Blood-Brain Barrier as a Pathogenic Mechanism for Central Nervous System Tuberculosis. J Infect Dis 193:1287–1295. doi: 10.1086/502631

40. Thwaites GE, van Toorn R, Schoeman J (2013) Tuberculous meningitis: more questions, still too few answers. Lancet Neurol 12:999–1010. doi: 10.1016/S1474-4422(13)70168-6

41. Sütlaş PN, Unal A, Forta H, et al (2003) Tuberculous meningitis in adults: review of 61 cases. Infection 31:387–391. doi: 10.1007/s15010-003-3179-1

42. Thwaites GE, Bang ND, Dung NH, et al (2005) The Influence of HIV Infection on Clinical Presentation, Response to Treatment, and Outcome in Adults with Tuberculous Meningitis. J Infect Dis 192:2134–2141. doi: 10.1086/498220

43. Farinha NJ, Razali KA, Holzel H, et al (2000) Tuberculosis of the central nervous system in children: a 20-year survey. J Infect 41:61–68. doi: 10.1053/jinf.2000.0692

44. Yaramiş A, Gurkan F, Elevli M, et al (1998) Central nervous system tuberculosis in children: a review of 214 cases. Pediatrics 102:E49.

45. Thwaites GE, Nguyen DB, Nguyen HD, et al (2004) Dexamethasone for the treatment of tuberculous meningitis in adolescents and adults. N Engl J Med 351:1741–1751. doi: 10.1056/NEJMoa040573

46. Thwaites G, Chau T, Stepniewska K, et al (2002) Diagnosis of adult tuberculous meningitis by use of clinical and laboratory features. The Lancet 360:1287–1292. doi: 10.1016/S0140-6736(02)11318-3

47. Sunbul M, Atilla A, Esen S, et al (2005) Thwaites' diagnostic scoring and the prediction of tuberculous meningitis. Med Princ Pract Int J Kuwait Univ Health Sci Cent 14:151–154. doi: 10.1159/000084631

48. Török ME, Nghia HDT, Chau TTH, et al (2007) Validation of a diagnostic algorithm for adult tuberculous meningitis. Am J Trop Med Hyg 77:555–559.

49. Vibha D, Bhatia R, Prasad K, et al (2012) Validation of diagnostic algorithm to differentiate between tuberculous meningitis and acute bacterial meningitis. Clin Neurol Neurosurg 114:639–644. doi: 10.1016/j.clineuro.2011.12.033

50. Checkley AM, Njalale Y, Scarborough M, Zjilstra EE (2008) Sensitivity and specificity of an index for the diagnosis of TB meningitis in patients in an urban teaching hospital in Malawi. Trop Med Int Health TM IH 13:1042–1046. doi: 10.1111/j.1365-3156.2008.02109.x

51. Stewart SM (1953) The Bacteriological Diagnosis of Tuberculous Meningitis. J Clin Pathol 6:241–242. doi: 10.1136/jcp.6.3.241

52. Thwaites G, Fisher M, Hemingway C, et al (2009) British Infection Society guidelines for the diagnosis and treatment of tuberculosis of the central nervous system in adults and children. J Infect 59:167–187. doi: 10.1016/j.jinf.2009.06.011

53. Tho DQ, Török ME, Yen NTB, et al (2012) Influence of Antituberculosis Drug Resistance and Mycobacterium tuberculosis Lineage on Outcome in HIV-Associated Tuberculous Meningitis. Antimicrob Agents Chemother 56:3074–3079. doi: 10.1128/AAC.00319-12

54. Scott LE, McCarthy K, Gous N, et al (2011) Comparison of Xpert MTB/RIF with Other Nucleic Acid Technologies for Diagnosing Pulmonary Tuberculosis in a High HIV Prevalence Setting: A Prospective Study. PLoS Med 8:e1001061. doi: 10.1371/journal.pmed.1001061

55. Hillemann D, Rüsch-Gerdes S, Boehme C, Richter E (2011) Rapid molecular detection of extrapulmonary tuberculosis by the automated GeneXpert MTB/RIF system. J Clin Microbiol 49:1202–1205. doi: 10.1128/JCM.02268-10

56. CDC | TB | Fact Sheets | Interferon-Gamma Release Assays (IGRAs). http://www.cdc.gov/tb/publications/factsheets/testing/igra.htm. Accessed 25 Mar 2015

57. Patel VB, Singh R, Connolly C, et al (2010) Comparison of a clinical prediction rule and a LAM antigen-detection assay for the rapid diagnosis of TBM in a high HIV prevalence setting. PloS One 5:e15664. doi: 10.1371/journal.pone.0015664

58. Kim S-H, Cho O-H, Park S-J, et al (2010) Rapid diagnosis of tuberculous meningitis by T cell-based assays on peripheral blood and cerebrospinal fluid mononuclear cells. Clin Infect Dis Off Publ Infect Dis Soc Am 50:1349–1358. doi: 10.1086/652142

59. Schutte C-M, Ungerer J p. j., du Plessis H, van der Meyden C h. (2001) Significance of cerebrospinal fluid adenosine deaminase isoenzymes in tuberculous (TB) meningitis. J Clin Lab Anal 15:236–238. doi: 10.1002/jcla.1033

60. Kashyap RS, Kainthla RP, Mudaliar AV, et al (2006) Cerebrospinal fluid adenosine deaminase activity: A complimentary tool in the early diagnosis of tuberculous meningitis. Cerebrospinal Fluid Res 3:5. doi: 10.1186/1743-8454-3-5

61. Chander A, Shrestha CD (2013) Cerebrospinal fluid adenosine deaminase levels as a diagnostic marker in tuberculous meningitis in adult Nepalese patients. Asian Pac J Trop Dis 3:16–19. doi: 10.1016/S2222-1808(13)60004-3

62. Choi S-H, Kim YS, Bae I-G, et al (2002) The possible role of cerebrospinal fluid adenosine deaminase activity in the diagnosis of tuberculous meningitis in adults. Clin Neurol Neurosurg 104:10–15.

63. Gupta BK, Bharat A, Debapriya B, Baruah H (2010) Adenosine Deaminase Levels in CSF of Tuberculous Meningitis Patients. J Clin Med Res 2:220–224. doi: 10.4021/jocmr429w

64. Malan C, Donald PR, Golden M, Taljaard JJ (1984) Adenosine deaminase levels in cerebrospinal fluid in the diagnosis of tuberculous meningitis. J Trop Med Hyg 87:33–40.

65. Mishra OP, Loiwal V, Ali Z, et al (1996) Cerebrospinal fluid adenosine deaminase activity for the diagnosis of tuberculous meningitis in children. J Trop Pediatr 42:129–132.

66. Corral I, Quereda C, Navas E, et al (2004) Adenosine deaminase activity in cerebrospinal fluid of HIV-infected patients: limited value for diagnosis of tuberculous meningitis. Eur J Clin Microbiol Infect Dis 23:471–476. doi: 10.1007/s10096-004-1110-z

67. Andronikou S, Smith B, Hatherhill M, et al (2004) Definitive neuroradiological diagnostic features of tuberculous meningitis in children. Pediatr Radiol 34:876–885. doi: 10.1007/s00247-004-1237-1

68. Kumar R, Kohli N, Thavnani H, et al (1996) Value of CT scan in the diagnosis of meningitis. Indian Pediatr 33:465–468.

69. CDC | TB | Treatment for TB Disease. http://www.cdc.gov/tb/topic/treatment/tbdisease.htm. Accessed 27 Mar 2015

70. Nau R, Sörgel F, Eiffert H (2010) Penetration of Drugs through the Blood-Cerebrospinal Fluid/Blood-Brain Barrier for Treatment of Central Nervous System Infections. Clin Microbiol Rev 23:858–883. doi: 10.1128/CMR.00007-10

71. Nau R, Prange HW, Menck S, et al (1992) Penetration of rifampicin into the cerebrospinal fluid of adults with uninflamed meninges. J Antimicrob Chemother 29:719–724.

72. Thwaites GE, Lan NTN, Dung NH, et al (2005) Effect of antituberculosis drug resistance on response to treatment and outcome in adults with tuberculous meningitis. J Infect Dis 192:79–88. doi: 10.1086/430616

73. Simmons CP, Thwaites GE, Quyen NTH, et al (2005) The clinical benefit of adjunctive dexamethasone in tuberculous meningitis is not associated with measurable attenuation of peripheral or local immune responses. J Immunol Baltim Md 1950 175:579–590.

74. Thwaites GE, Macmullen-Price J, Tran THC, et al (2007) Serial MRI to determine the effect of dexamethasone on the cerebral pathology of tuberculous meningitis: an observational study. Lancet Neurol 6:230–236. doi: 10.1016/S1474-4422(07)70034-0

75. van Toorn R, du Plessis A-M, Schaaf HS, et al (2015) Clinicoradiologic response of neurologic tuberculous mass lesions in children treated with thalidomide. Pediatr Infect Dis J 34:214–218. doi: 10.1097/INF.0000000000000539

76. Sampaio EP, Sarno EN, Galilly R, et al (1991) Thalidomide selectively inhibits tumor necrosis factor alpha production by stimulated human monocytes. J Exp Med 173:699–703.

77. van Toorn R, Schaaf HS, Laubscher JA, et al (2014) Short intensified treatment in children with drug-susceptible tuberculous meningitis. Pediatr Infect Dis J 33:248–252. doi: 10.1097/INF.0000000000000065

78. Jucker M (2010) The benefits and limitations of animal models for translational research in neurodegenerative diseases. Nat Med 16:1210–1214. doi: 10.1038/nm.2224

79. Tsenova L, Bergtold A, Freedman VH, et al (1999) Tumor necrosis factor alpha is a determinant of pathogenesis and disease progression in mycobacterial infection in the central nervous system. Proc Natl Acad Sci U S A 96:5657–5662.

80. Tsenova L, Harbacheuski R, Moreira AL, et al (2006) Evaluation of the Mtb72F polyprotein vaccine in a rabbit model of tuberculous meningitis. Infect Immun 74:2392–2401. doi: 10.1128/IAI.74.4.2392-2401.2006

81. Tsenova L, Mangaliso B, Muller G, et al (2002) Use of IMiD3, a Thalidomide Analog, as an Adjunct to Therapy for Experimental Tuberculous Meningitis. Antimicrob Agents Chemother 46:1887–1895. doi: 10.1128/AAC.46.6.1887-1895.2002

82. Tsenova L, Sokol K, Victoria HF, Kaplan G (1998) A Combination of Thalidomide plus Antibiotics Protects Rabbits from Mycobacterial Meningitis-Associated Death. J Infect Dis 177:1563–1572. doi: 10.1086/515327

83. Tsenova L, Ellison E, Harbacheuski R, et al (2005) Virulence of selected Mycobacterium tuberculosis clinical isolates in the rabbit model of meningitis is dependent on phenolic glycolipid produced by the bacilli. J Infect Dis 192:98–106. doi: 10.1086/430614

84. Mazzolla R, Puliti M, Barluzzi R, et al (2002) Differential microbial clearance and immunoresponse of Balb/c (Nramp1 susceptible) and DBA2 (Nramp1 resistant) mice intracerebrally infected with Mycobacterium bovis BCG (BCG). FEMS Immunol Med Microbiol 32:149–158.

85. Well GTJ van, Wieland CW, Florquin S, et al (2007) A New Murine Model to Study the Pathogenesis of Tuberculous Meningitis. J Infect Dis 195:694–697. doi: 10.1086/511273

86. Santoriello C, Zon LI (2012) Hooked! Modeling human disease in zebrafish. J Clin Invest 122:2337–2343. doi: 10.1172/JCI60434

87. Motsinger-Reif AA, Antas PRZ, Oki NO, et al (2010) Polymorphisms in IL-1beta, vitamin D receptor Fok1, and Toll-like receptor 2 are associated with extrapulmonary tuberculosis. BMC Med Genet 11:37. doi: 10.1186/1471-2350-11-37

88. Hawn TR, Dunstan SJ, Thwaites GE, et al (2006) A polymorphism in Toll-interleukin 1 receptor domain containing adaptor protein is associated with susceptibility to meningeal tuberculosis. JInfectDis 194:1127–1134.

89. Texereau J, Chiche J-D, Taylor W, et al (2005) The importance of Toll-like receptor 2 polymorphisms in severe infections. Clin Infect Dis Off Publ Infect Dis Soc Am 41 Suppl 7:S408-415. doi: 10.1086/431990

90. Bafica A, Scanga CA, Feng CG, et al (2005) TLR9 regulates Th1 responses and cooperates with TLR2 in mediating optimal resistance to Mycobacterium tuberculosis. J Exp Med 202:1715–1724. doi: 10.1084/jem.20051782

91. Heldwein KA, Liang MD, Andresen TK, et al (2003) TLR2 and TLR4 serve distinct roles in the host immune response against Mycobacterium bovis BCG. JLeukocBiol 74:277–286.

92. Sugawara I, Yamada H, Mizuno S, et al (2003) Mycobacterial infection in MyD88-deficient mice. Microbiol Immunol 47:841–847.

93. El-Kebir M, van der Kuip M, van Furth AM, Kirschner DE (2013) Computational modeling of tuberculous meningitis reveals an important role for tumor necrosis factor-α. J Theor Biol 328:43–53. doi: 10.1016/j.jtbi.2013.03.008

94. Pfeffer K (2003) Biological functions of tumor necrosis factor cytokines and their receptors. Cytokine Growth Factor Rev 14:185–191. doi: 10.1016/S1359-6101(03)00022-4

95. Bernard C, Tedgui A (1992) Cytokine network and the vessel wall. Insights into septic shock pathogenesis. Eur Cytokine Netw 3:19–33.

96. Tureen J (1995) Effect of recombinant human tumor necrosis factor-alpha on cerebral oxygen uptake, cerebrospinal fluid lactate, and cerebral blood flow in the rabbit: role of nitric oxide. J Clin Invest 95:1086–1091.

97. Yoshizumi M, Perrella MA, Burnett JC, Lee ME (1993) Tumor necrosis factor downregulates an endothelial nitric oxide synthase mRNA by shortening its half-life. Circ Res 73:205–209. doi: 10.1161/01.RES.73.1.205

98. Tobin DM, Roca FJ, Oh SF, et al (2012) Host genotype-specific therapies can optimize the inflammatory response to mycobacterial infections. Cell 148:434–446. doi: 10.1016/j.cell.2011.12.023

99. Tobin DM, Vary JC, Ray JP, et al (2010) The lta4h locus modulates susceptibility to mycobacterial infection in zebrafish and humans. Cell 140:717–730. doi: 10.1016/j.cell.2010.02.013

100. Behr M, Schurr E, Gros P (2010) TB: screening for responses to a vile visitor. Cell 140:615–618. doi: 10.1016/j.cell.2010.02.030

101. Schlesinger L, Horwitz M (1991) Phenolic glycolipid-1 of Mycobacterium leprae binds complement component C3 in serum and mediates phagocytosis by human monocytes. J Exp Med 174:1031–1038.

102. Mastroianni CM, Lancella L, Mengoni F, et al (1998) Chemokine profiles in the cerebrospinal fluid (CSF) during the course of pyogenic and tuberculous meningitis. Clin Exp Immunol 114:210–214.

103. Nagesh Babu G, Kumar A, Kalita J, Misra UK (2008) Proinflammatory cytokine levels in the serum and cerebrospinal fluid of tuberculous meningitis patients. Neurosci Lett 436:48–51. doi: 10.1016/j.neulet.2008.02.060

104. Rock RB, Gekker G, Hu S, et al (2004) Role of microglia in central nervous system infections. Clin Microbiol Rev 17:942–964, table of contents. doi: 10.1128/CMR.17.4.942-964.2004

105. Berenguer J, Moreno S, Laguna F, et al (1992) Tuberculous Meningitis in Patients Infected with the Human Immunodeficiency Virus. N Engl J Med 326:668–672. doi: 10.1056/NEJM199203053261004

106. Garg RK, Sinha MK (2010) Tuberculous meningitis in patients infected with human immunodeficiency virus. J Neurol 258:3–13. doi: 10.1007/s00415-010-5744-8

107. Johnson MD, Decker CF (2006) Tuberculosis and HIV Infection. DisMon 52:420–427.

108. Selwyn PA, Hartel D, Lewis VA, et al (1989) A Prospective Study of the Risk of Tuberculosis among Intravenous Drug Users with Human Immunodeficiency Virus Infection. N Engl J Med 320:545–550. doi: 10.1056/NEJM198903023200901

109. Caws M, Thwaites G, Stepniewska K, et al (2006) Beijing genotype of Mycobacterium tuberculosis is significantly associated with human immunodeficiency virus infection and multidrug resistance in cases of tuberculous meningitis. J Clin Microbiol 44:3934–3939. doi: 10.1128/JCM.01181-06

110. Marais S, Pepper DJ, Marais BJ, Török ME (2010) HIV-associated tuberculous meningitis – diagnostic and therapeutic challenges. Tuberculosis 90:367–374. doi: 10.1016/j.tube.2010.08.006

111. Torok ME, Chau TTH, Mai PP, et al (2008) Clinical and Microbiological Features of HIV-Associated Tuberculous Meningitis in Vietnamese Adults. PLoS ONE 3:e1772. doi: 10.1371/journal.pone.0001772

112. Patel NR, Swan K, Li X, et al (2009) Impaired M. tuberculosis-mediated apoptosis in alveolar macrophages from HIV+ persons: potential role of IL-10 and BCL-3. J Leukoc Biol 86:53–60. doi: 10.1189/jlb.0908574

113. Patel NR, Zhu J, Tachado SD, et al (2007) HIV impairs TNF-alpha mediated macrophage apoptotic response to Mycobacterium tuberculosis. J Immunol Baltim Md 1950 179:6973–6980.

114. Möller M, de Wit E, Hoal EG (2010) Past, present and future directions in human genetic susceptibility to tuberculosis. FEMS ImmunolMedMicrobiol 58:3–26.

115. Bellamy R, Ruwende C, Corrah T, et al (1998) Variations in the NRAMP1 Gene and Susceptibility to Tuberculosis in West Africans. N Engl J Med 338:640–644. doi: 10.1056/NEJM199803053381002

116. Caws M, Thwaites G, Dunstan S, et al (2008) The influence of host and bacterial genotype on the development of disseminated disease with Mycobacterium tuberculosis. PLoSPathog 4:e1000034.

117. Dissanayeke SR, Levin S, Pienaar S, et al (2009) Polymorphic variation in TIRAP is not associated with susceptibility to childhood TB but may determine susceptibility to TBM in some ethnic groups. PLoSONE 4:e6698.

118. Thuong NT, Hawn TR, Thwaites GE, et al (2007) A polymorphism in human TLR2 is associated with increased susceptibility to tuberculous meningitis. Genes Immun 8:422–428.

119. Cambier CJ, Takaki KK, Larson RP, et al (2014) Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. Nature 505:218–222. doi: 10.1038/nature12799

120. Reed MB, Domenech P, Manca C, et al (2004) A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. Nature 431:84–87.

121. Castiblanco J, Varela D-C, Castaño-Rodríguez N, et al (2008) TIRAP (MAL) S180L polymorphism is a common protective factor against developing tuberculosis and systemic lupus erythematosus. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 8:541–544. doi: 10.1016/j.meegid.2008.03.001

122. Khor CC, Chapman SJ, Vannberg FO, et al (2007) A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. Nat Genet 39:523–528. doi: 10.1038/ng1976

123. Fernando SL, Saunders BM, Sluyter R, et al (2007) A polymorphism in the P2X7 gene increases susceptibility to extrapulmonary tuberculosis. AmJRespirCrit Care Med 175:360–366.

124. Kusner DJ, Adams J (2000) ATP-Induced Killing of Virulent Mycobacterium tuberculosis Within Human Macrophages Requires Phospholipase D. J Immunol 164:379–388. doi: 10.4049/jimmunol.164.1.379

125. Kusner DJ, Barton JA (2001) ATP Stimulates Human Macrophages to Kill Intracellular Virulent Mycobacterium tuberculosis Via Calcium-Dependent Phagosome-Lysosome Fusion. J Immunol 167:3308–3315. doi: 10.4049/jimmunol.167.6.3308

126. Ferrari D, Los M, Bauer MK, et al (1999) P2Z purinoreceptor ligation induces activation of caspases with distinct roles in apoptotic and necrotic alterations of cell death. FEBS Lett 447:71–75.

127. Humphreys BD, Rice J, Kertesy SB, Dubyak GR (2000) Stress-activated protein kinase/JNK activation and apoptotic induction by the macrophage P2X7 nucleotide receptor. J Biol Chem 275:26792–26798. doi: 10.1074/jbc.M002770200

128. Wiley JS, Gargett CE, Zhang W, et al (1998) Partial agonists and antagonists reveal a second permeability state of human lymphocyte P2Z/P2X7 channel. Am J Physiol 275:C1224-1231.

129. Fairbairn IP, Stober CB, Kumararatne DS, Lammas DA (2001) ATP-Mediated Killing of Intracellular Mycobacteria by Macrophages Is a P2X7-Dependent Process Inducing Bacterial Death by Phagosome-Lysosome Fusion. J Immunol 167:3300–3307. doi: 10.4049/jimmunol.167.6.3300

130. Gu BJ, Zhang W, Worthington RA, et al (2001) A Glu-496 to Ala polymorphism leads to loss of function of the human P2X7 receptor. J Biol Chem 276:11135–11142. doi: 10.1074/jbc.M010353200

131. Wiley JS, Dao-Ung L-P, Li C, et al (2003) An Ile-568 to Asn polymorphism prevents normal trafficking and function of the human P2X7 receptor. J Biol Chem 278:17108–17113. doi: 10.1074/jbc.M212759200

132. Gu BJ, Sluyter R, Skarratt KK, et al (2004) An Arg307 to Gln polymorphism within the ATP-binding site causes loss of function of the human P2X7 receptor. J Biol Chem 279:31287–31295. doi: 10.1074/jbc.M313902200

133. Skarratt KK, Fuller SJ, Sluyter R, et al (2005) A 5' intronic splice site polymorphism leads to a null allele of the P2X7 gene in 1-2% of the Caucasian population. FEBS Lett 579:2675–2678. doi: 10.1016/j.febslet.2005.03.091

134. Fernando SL, Saunders BM, Sluyter R, et al (2005) Gene Dosage Determines the Negative Effects of Polymorphic Alleles of the P2X7 Receptor on Adenosine Triphosphate–Mediated Killing of Mycobacteria by Human Macrophages. J Infect Dis 192:149–155. doi: 10.1086/430622

135. Saunders BM, Fernando SL, Sluyter R, et al (2003) A Loss-of-Function Polymorphism in the Human P2X7 Receptor Abolishes ATP-Mediated Killing of Mycobacteria. J Immunol 171:5442–5446. doi: 10.4049/jimmunol.171.10.5442

136. Li CM, Campbell SJ, Kumararatne DS, et al (2002) Association of a polymorphism in the P2X7 gene with tuberculosis in a Gambian population. J Infect Dis 186:1458–1462. doi: 10.1086/344351

137. Wu G, Zhao M, Gu X, et al (2014) The effect of P2X7 receptor 1513 polymorphism on susceptibility to tuberculosis: A meta-analysis. Infect Genet Evol 24:82–91. doi: 10.1016/j.meegid.2014.03.006

138. Ozdemir FA, Erol D, Konar V, et al (2014) Lack of association of 1513 A/C polymorphism in P2X7 gene with susceptibility to pulmonary and extrapulmonary tuberculosis. Tüberküloz Ve Toraks 62:7–11.

139. Lee J, Kornfeld H (2010) Interferon-γ Regulates the Death of M. tuberculosis-Infected Macrophages. J Cell Death 3:1–11.

140. Rossouw M, Nel HJ, Cooke GS, et al (2003) Association between tuberculosis and a polymorphic NFkappaB binding site in the interferon gamma gene. Lancet 361:1871–1872. doi: 10.1016/S0140-6736(03)13491-5

141. Pacheco AG, Cardoso CC, Moraes MO (2008) IFNG +874T/A, IL10 -1082G/A and TNF -308G/A polymorphisms in association with tuberculosis susceptibility: a meta-analysis study. Hum Genet 123:477–484. doi: 10.1007/s00439-008-0497-5

142. Shen C, Jiao W-W, Feng W-X, et al (2013) IFNG polymorphisms are associated with tuberculosis in Han Chinese pediatric female population. Mol Biol Rep 40:5477–5482. doi: 10.1007/s11033-013-2647-7

143. Qi H, Sun L, Jin Y-Q, et al (2014) rs2243268 and rs2243274 of Interleukin-4 (IL-4) gene are associated with reduced risk for extrapulmonary and severe tuberculosis in Chinese Han children. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 23:121–128. doi: 10.1016/j.meegid.2014.01.031

144. Wurtz O, Bajénoff M, Guerder S (2004) IL-4-mediated inhibition of IFN-gamma production by CD4+ T cells proceeds by several developmentally regulated mechanisms. Int Immunol 16:501–508.

145. Faksri K, Drobniewski F, Nikolayevskyy V, et al (2011) Epidemiological trends and clinical comparisons of Mycobacterium tuberculosis lineages in Thai TB meningitis. Tuberc Edinb Scotl 91:594–600. doi: 10.1016/j.tube.2011.08.005

146. Kong Y, Cave MD, Zhang L, et al (2007) Association between Mycobacterium tuberculosis Beijing/W Lineage Strain Infection and Extrathoracic Tuberculosis: Insights from Epidemiologic and Clinical Characterization of the Three Principal Genetic Groups of M. tuberculosis Clinical Isolates. J Clin Microbiol 45:409–414. doi: 10.1128/JCM.01459-06

147. Yorsangsukkamol J, Chaiprasert A, Prammananan T, et al (2009) Molecular analysis of Mycobacterium tuberculosis from tuberculous meningitis patients in Thailand. Tuberculosis 89:304–309. doi: 10.1016/j.tube.2009.05.001

148. Nicol MP, Sola C, February B, et al (2005) Distribution of strain families of Mycobacterium tuberculosis causing pulmonary and extrapulmonary disease in hospitalized children in Cape Town, South Africa. J Clin Microbiol 43:5779–5781. doi: 10.1128/JCM.43.11.5779-5781.2005

149. Quy HT, Cobelens FGJ, Lan NTN, et al (2006) Treatment outcomes by drug resistance and HIV status among tuberculosis patients in Ho Chi Minh City, Vietnam. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis 10:45–51.

150. Thwaites GE, Chau TTH, Caws M, et al (2002) Isoniazid resistance, mycobacterial genotype and outcome in Vietnamese adults with tuberculous meningitis. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis 6:865–871.

151. Karstaedt AS, Valtchanova S, Barriere R, Crewe-Brown HH (1998) Tuberculous meningitis in South African urban adults. QJM Mon J Assoc Physicians 91:743–747.

152. Health Systems Trust (2009) Incidence Of TB In The Provinces Of South Africa. In: http://www.hst.org.za/healthstats/16/data. http://www.hst.org.za/healthstats/16/data.

153. Microsoft Word - Census in Brief Ros updates 28 Oct 2012.doc - Census_2011_Census_in_brief.pdf. http://www.statssa.gov.za/census/census_2011/census_products/Census_2011_Census_in_brief.pdf. Accessed 22 Jul 2016

154. de Wit E, Delport W, Rugamika CE, et al (2010) Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. HumGenet 128:145–153.

155. Healy DG (2006) Case-control studies in the genomic era: a clinician's guide. Lancet Neurol 5:701–707.

156. Munch Z, Van Lill SWP, Booysen CN, et al (2003) Tuberculosis transmission patterns in a high-incidence area: a spatial analysis. Int J Tuberc Lung Dis 7:271–277.

157. Beyers N, Gie RP, Zietsman HL, et al (1996) The use of a geographical information system (GIS) to evaluate the distribution of tuberculosis in a high-incidence community. SAfrMedJ 86:40–1, 44.

158. Daya M, van der Merwe L, van Helden PD, et al (2015) Investigating the Role of Gene-Gene Interactions in TB Susceptibility. PLoS ONE. doi: 10.1371/journal.pone.0123970

159. Ng SB, Buckingham KJ, Lee C, et al (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42:30–35. doi: 10.1038/ng.499

160. Sakharkar MK, Chow VTK, Kangueane P (2004) Distributions of exons and introns in the human genome. In Silico Biol 4:387–393.

161. Gnirke A, Melnikov A, Maguire J, et al (2009) Solution Hybrid Selection with Ultra-long Oligonucleotides for Massively Parallel Targeted Sequencing. Nat Biotechnol 27:182–189. doi: 10.1038/nbt.1523

162. Christodoulou K, Wiskin AE, Gibson J, et al (2013) Next generation exome sequencing of paediatric inflammatory bowel disease patients identifies rare and novel variants in candidate genes. Gut 62:977–984. doi: 10.1136/gutjnl-2011-301833

163. Emond MJ, Louie T, Emerson J, et al (2015) Exome Sequencing of Phenotypic Extremes Identifies CAV2 and TMC6 as Interacting Modifiers of Chronic Pseudomonas aeruginosa Infection in Cystic Fibrosis. PLoS Genet 11:e1005273. doi: 10.1371/journal.pgen.1005273

164. Emond MJ, Louie T, Emerson J, et al (2012) Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nat Genet 44:886–889. doi: 10.1038/ng.2344

165. Bamshad MJ, Ng SB, Bigham AW, et al (2011) Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12:745–755. doi: 10.1038/nrg3031

166. Schork NJ, Murray SS, Frazer KA, Topol EJ (2009) Common vs. Rare Allele Hypotheses for Complex Diseases. Curr Opin Genet Dev 19:212–219. doi: 10.1016/j.gde.2009.04.010

167. Li A, Meyre D (2013) Challenges in reproducibility of genetic association studies: lessons learned from the obesity field. Int J Obes 2005 37:559–567. doi: 10.1038/ijo.2012.82

168. Auer PL, Lettre G (2015) Rare variant association studies: considerations, challenges and opportunities. Genome Med. doi: 10.1186/s13073-015-0138-2

169. Bansal V, Libiger O, Torkamani A, Schork NJ (2010) Statistical Analysis Strategies for Association Studies Involving Rare Variants. Nat Rev Genet 11:773–785. doi: 10.1038/nrg2867

170. Huang H, Chanda P, Alonso A, et al (2011) Gene-Based Tests of Association. PLOS Genet 7:e1002177. doi: 10.1371/journal.pgen.1002177

171. Lee S, Abecasis GR, Boehnke M, Lin X (2014) Rare-Variant Association Analysis: Study Designs and Statistical Tests. Am J Hum Genet 95:5–23. doi: 10.1016/j.ajhg.2014.06.009

172. Neale BM, Sham PC (2004) The Future of Association Studies: Gene-Based Analysis and Replication. Am J Hum Genet 75:353–362. doi: 10.1086/423901

173. Morris AP, Zeggini E (2010) An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol 34:188–193. doi: 10.1002/gepi.20450

174. Pritchard JK (2001) Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 69:124–137. doi: 10.1086/321272

175. Bobadilla JL, Macek M, Fine JP, Farrell PM (2002) Cystic fibrosis: a worldwide analysis of CFTR mutations--correlation with incidence data and application to screening. Hum Mutat 19:575–606. doi: 10.1002/humu.10041

176. Ma X, Liu Y, Gowen BB, et al (2007) Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. PLoSONE 2:e1318.

177. Morgenthaler S, Thilly WG (2007) A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: A cohort allelic sums test (CAST). Mutat Res Mol Mech Mutagen 615:28–56. doi: 10.1016/j.mrfmmm.2006.09.003

178. Neale BM, Rivas MA, Voight BF, et al (2011) Testing for an Unusual Distribution of Rare Variants. PLOS Genet 7:e1001322. doi: 10.1371/journal.pgen.1001322

179. Wu MC, Lee S, Cai T, et al (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89:82–93. doi: 10.1016/j.ajhg.2011.05.029

180. Lee S, Wu MC, Lin X (2012) Optimal tests for rare variant effects in sequencing association studies. Biostat Oxf Engl 13:762–775. doi: 10.1093/biostatistics/kxs014

181. Ionita-Laza I, Lee S, Makarov V, et al (2013) Sequence kernel association tests for the combined effect of rare and common variants. Am J Hum Genet 92:841–853. doi: 10.1016/j.ajhg.2013.04.015

182. Metzker ML (2010) Sequencing technologies — the next generation. Nat Rev Genet 11:31–46. doi: 10.1038/nrg2626

183. Kent WJ, Sugnet CW, Furey TS, et al (2002) The Human Genome Browser at UCSC. Genome Res 12:996–1006. doi: 10.1101/gr.229102

184. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma Oxf Engl 25:1754–1760. doi: 10.1093/bioinformatics/btp324

185. Robinson JT, Thorvaldsdóttir H, Winckler W, et al (2011) Integrative genomics viewer. Nat Biotechnol 29:24–26. doi: 10.1038/nbt.1754

186. Koboldt DC, Ding L, Mardis ER, Wilson RK (2010) Challenges of sequencing human genomes. Brief Bioinform bbq016. doi: 10.1093/bib/bbq016

187. Barnett D, Garrison E, Quinlan A, et al (2011) BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics btr174. doi: 10.1093/bioinformatics/btr174

188. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 15 Mar 2016

189. Metzker ML (2010) Sequencing technologies — the next generation. Nat Rev Genet 11:31–46. doi: 10.1038/nrg2626

190. Gorlov IP, Gorlova OY, Sunyaev SR, et al (2008) Shifting Paradigm of Association Studies: Value of Rare Single-Nucleotide Polymorphisms. Am J Hum Genet 82:100–112. doi: 10.1016/j.ajhg.2007.09.006

191. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164. doi: 10.1093/nar/gkq603

192. Consortium T 1000 GP (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65. doi: 10.1038/nature11632

193. Exome Variant Server. http://evs.gs.washington.edu/EVS/. Accessed 15 Mar 2016

194. ExAC Browser. http://exac.broadinstitute.org/. Accessed 12 Jul 2016

195. Ng PC, Henikoff S (2002) Accounting for Human Polymorphisms Predicted to Affect Protein Function. Genome Res 12:436–446. doi: 10.1101/gr.212802

196. Cirulli ET, Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet 11:415–425. doi: 10.1038/nrg2779

197. PHAST: Home. http://compgen.cshl.edu/phast/. Accessed 15 Mar 2016

198. Sidow Lab - GERP. http://mendel.stanford.edu/SidowLab/downloads/gerp/. Accessed 15 Mar 2016

199. Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. doi: 10.1093/bioinformatics/btr330

200. Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. doi: 10.1086/519795

201. R-programming environment.

202. Lee S, Emond MJ, Bamshad MJ, et al (2012) Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 91:224–237. doi: 10.1016/j.ajhg.2012.06.007

203. Albert A, Anderson JA (1984) On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. Biometrika 71:1–10. doi: 10.2307/2336390

204. Farinha NJ, Razali KA, Holzel H, et al (2000) Tuberculosis of the Central Nervous System in Children: a 20-Year Survey. J Infect 41:61–68. doi: 10.1053/jinf.2000.0692

205. Lee S, Emond MJ, Bamshad MJ, et al (2012) Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies. Am J Hum Genet 91:224–237. doi: 10.1016/j.ajhg.2012.06.007

206. Ingenuity Pathway Analysis. Qiagen, Redwood City, CA

207. Sims D, Sudbery I, Ilott NE, et al (2014) Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 15:121–132. doi: 10.1038/nrg3642

208. Choi J-M, Kim S-H, Shin J-H, et al (2008) Transduction of the cytoplasmic domain of CTLA-4 inhibits TcR-specific activation signals and prevents collagen-induced arthritis. Proc Natl Acad Sci U S A 105:19875–19880. doi: 10.1073/pnas.0805198105

209. Bonini SA, Ferrari-Toninelli G, Uberti D, et al (2011) Nuclear Factor κB-Dependent Neurite Remodeling Is Mediated by Notch Pathway. J Neurosci 31:11697–11705. doi: 10.1523/JNEUROSCI.1113-11.2011

210. Chauhan VS, Sterka DG, Furr SR, et al (2009) NOD2 plays an important role in the inflammatory responses of microglia and astrocytes to bacterial CNS pathogens. Glia 57:414–423. doi: 10.1002/glia.20770

211. Cui B, Zhang S, Chen L, et al (2013) Targeting ROR1 inhibits epithelial-mesenchymal transition and metastasis. Cancer Res 73:3649–3660. doi: 10.1158/0008-5472.CAN-12-3832

212. Fan H, Zhao Z, Quan Y, et al (2007) DNA methyltransferase 1 knockdown induces silenced CDH1 gene reexpression by demethylation of methylated CpG in hepatocellular carcinoma cell line SMMC-7721. Eur J Gastroenterol Hepatol 19:952–961. doi: 10.1097/MEG.0b013e3282c3a89e

213. Kikuta Y, Miyauchi Y, Kusunose E, Kusunose M (1999) Expression and Molecular Cloning of Human Liver Leukotriene B4 omega-Hydroxylase (CYP4F2) Gene. DNA Cell Biol 18:723–730. doi: 10.1089/104454999315006

214. Jin R, Koop DR, Raucy JL, Lasker JM (1998) Role of Human CYP4F2 in Hepatic Catabolism of the Proinflammatory Agent Leukotriene B4. Arch Biochem Biophys 359:89–98. doi: 10.1006/abbi.1998.0880

215. Johnson AL, Edson KZ, Totah RA, Rettie AE (2015) Cytochrome P450 ω-Hydroxylases in Inflammation and Cancer. Adv Pharmacol San Diego Calif 74:223–262. doi: 10.1016/bs.apha.2015.05.002

216. McLaren W, Pritchard B, Rios D, et al (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 26:2069–2070. doi: 10.1093/bioinformatics/btq330

217. Costea IC, Mack DR, Israel DM, et al (2008) M2069 Association of the CYP4F2 Gene, a Regulator of Leukotriene B4 (LTB4) with Risk for Crohn's Disease in Children. Gastroenterology 134:A-462. doi: 10.1016/S0016-5085(08)62162-9

218. Liu X, Chauhan VS, Young AB, Marriott I (2010) NOD2 mediates inflammatory responses of primary murine glia to Streptococcus pneumoniae. Glia 58:839–847. doi: 10.1002/glia.20968

219. Girardin SE, Boneca IG, Viala J, et al (2003) Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. JBiolChem 278:8869–8872.

220. Möller M, Nebel A, Kwiatkowski R, et al (2007) Host susceptibility to tuberculosis: CARD15 polymorphisms in a South African population. Mol Cell Probes 21:148–151. doi: 10.1016/j.mcp.2006.10.001

221. Schulze TG, Zhang K, Chen Y-S, et al (2004) Defining haplotype blocks and tag single-nucleotide polymorphisms in the human genome. Hum Mol Genet 13:335–342. doi: 10.1093/hmg/ddh035

222. Tsang WY, Bossard C, Khanna H, et al (2008) CP110 Suppresses Primary Cilia Formation through its Interaction with CEP290, a Protein Deficient in Human Ciliary Disease. Dev Cell 15:187–197. doi: 10.1016/j.devcel.2008.07.004

223. Saunders NR, Liddelow SA, Dziegielewska KM (2012) Barrier Mechanisms in the Developing Brain. Front Pharmacol. doi: 10.3389/fphar.2012.00046

224. LaMarre J, Wollenberg GK, Gonias SL, Hayes MA (1991) Cytokine binding and clearance properties of proteinase-activated alpha 2-macroglobulins. Lab Investig J Tech Methods Pathol 65:3–14.

225. Kuribayashi T, Tomizawa M, Seita T, et al (2011) Relationship between production of acute-phase proteins and strength of inflammatory stimulation in rats. Lab Anim 45:215–218. doi: 10.1258/la.2011.010112

226. Kanoh Y, Ohara T, Akahoshi T (2011) Acute inflammatory biomarkers in cerebrospinal fluid as indicators of blood cerebrospinal fluid barrier damage in Japanese subjects with infectious meningitis. Clin Lab 57:37–46.

227. Jain S, Gautam V, Naseem S (2011) Acute-phase proteins: As diagnostic tool. J Pharm Bioallied Sci 3:118–127. doi: 10.4103/0975-7406.76489

228. Kumar NP, Anuradha R, Andrade BB, et al (2013) Circulating Biomarkers of Pulmonary and Extrapulmonary Tuberculosis in Children. Clin Vaccine Immunol 20:704–711. doi: 10.1128/CVI.00038-13

229. Bapat PR, Satav AR, Husain AA, et al (2015) Differential Levels of Alpha-2-Macroglobulin, Haptoglobin and Sero-Transferrin as Adjunct Markers for TB Diagnosis and Disease Progression in the Malnourished Tribal Population of Melghat, India. PLOS ONE 10:e0133928. doi: 10.1371/journal.pone.0133928

230. Creemers EEJM, Cleutjens JPM, Smits JFM, Daemen MJAP (2001) Matrix Metalloproteinase Inhibition After Myocardial Infarction A New Approach to Prevent Heart Failure? Circ Res 89:201–210. doi: 10.1161/hh1501.094396

231. Leib SL, Leppert D, Clements J, Täuber MG (2000) Matrix Metalloproteinases Contribute to Brain Damage in Experimental Pneumococcal Meningitis. Infect Immun 68:615–620. doi: 10.1128/IAI.68.2.615-620.2000

232. Lakhan SE, Kirchgessner A, Tepper D, Leonard A (2013) Matrix Metalloproteinases and Blood-Brain Barrier Disruption in Acute Ischemic Stroke. Front Neurol. doi: 10.3389/fneur.2013.00032

233. Malipiero U, Koedel U, Pfister H-W, et al (2006) TGFbeta receptor II gene deletion in leucocytes prevents cerebral vasculitis in bacterial meningitis. Brain J Neurol 129:2404–2415. doi: 10.1093/brain/awl192

234. Ferretti E, Pistoia V, Corcione A, et al (2014) Role of Fractalkine/CX3CL1 and Its Receptor in the Pathogenesis of Inflammatory and Malignant Diseases with Emphasis on B Cell Malignancies, Role of Fractalkine/CX3CL1 and Its Receptor in the Pathogenesis of Inflammatory and Malignant Diseases with Emphasis on B Cell Malignancies. Mediat Inflamm Mediat Inflamm 2014, 2014:e480941. doi: 10.1155/2014/480941, 10.1155/2014/480941

235. Jones BA, Beamer M, Ahmed S (2010) Fractalkine/CX3CL1: A Potential New Target for Inflammatory Diseases. Mol Interv 10:263–270. doi: 10.1124/mi.10.5.3

236. Wolf Y, Yona S, Kim K-W, Jung S (2013) Microglia, seen from the CX3CR1 angle. Front Cell Neurosci. doi: 10.3389/fncel.2013.00026

237. Liu Y, Wu X-M, Luo Q-Q, et al (2015) CX3CL1/CX3CR1-mediated microglia activation plays a detrimental role in ischemic mice brain via p38MAPK/PKC pathway. J Cereb Blood Flow Metab Off J Int Soc Cereb Blood Flow Metab 35:1623–1631. doi: 10.1038/jcbfm.2015.97

238. Hall JD, Kurtz SL, Rigel NW, et al (2009) The impact of chemokine receptor CX3CR1 deficiency during respiratory infections with Mycobacterium tuberculosis or Francisella tularensis. Clin Exp Immunol 156:278–284. doi: 10.1111/j.1365-2249.2009.03882.x

239. Chen S, Luo D, Streit WJ, Harrison JK (2002) TGF-beta1 upregulates CX3CR1 expression and inhibits fractalkine-stimulated signaling in rat microglia. J Neuroimmunol 133:46–55.

240. Schenck A, Bardoni B, Moro A, et al (2001) A highly conserved protein family interacting with the fragile X mental retardation protein (FMRP) and displaying selective interactions with FMRP-related proteins FXR1P and FXR2P. Proc Natl Acad Sci U S A 98:8844–8849. doi: 10.1073/pnas.151231598

241. Wray S (2010) From Nose to Brain: Development of Gonadotropin-releasing hormone −1 Neurons. J Neuroendocrinol 22:743–753. doi: 10.1111/j.1365-2826.2010.02034.x

242. Shi L, Chang X, Zhang P, et al (2013) The functional genetic link of NLGN4X knockdown and neurodevelopment in neural stem cells. Hum Mol Genet 22:3749–3760. doi: 10.1093/hmg/ddt226

243. Chen H-J, Lin C-M, Lin C-S, et al (2006) The role of microtubule actin cross-linking factor 1 (MACF1) in the Wnt signaling pathway. Genes Dev 20:1933–1945. doi: 10.1101/gad.1411206

244. Bowdish DME, Sakamoto K, Kim M-J, et al (2009) MARCO, TLR2, and CD14 Are Required for Macrophage Cytokine Responses to Mycobacterial Trehalose Dimycolate and Mycobacterium tuberculosis. PLoS Pathog. doi: 10.1371/journal.ppat.1000474

245. Braun BJ, Slowik A, Leib SL, et al (2011) The formyl peptide receptor like-1 and scavenger receptor MARCO are involved in glial cell activation in bacterial meningitis. J Neuroinflammation 8:11. doi: 10.1186/1742-2094-8-11

246. Ito S, Naito M, Kobayashi Y, et al (1999) Roles of a macrophage receptor with collagenous structure (MARCO) in host defense and heterogeneity of splenic marginal zone macrophages. Arch Histol Cytol 62:83–95.

247. Ma M-J, Wang H-B, Li H, et al (2011) Genetic variants in MARCO are associated with the susceptibility to pulmonary tuberculosis in Chinese Han population. PloS One 6:e24069. doi: 10.1371/journal.pone.0024069

248. Bowdish DM, Sakamoto K, Lack NA, et al (2013) Genetic variants of MARCO are associated with susceptibility to pulmonary tuberculosis in a Gambian population. BMC Med Genet 14:47. doi: 10.1186/1471-2350-14-47

249. Paganoni S, Ferreira A (2005) Neurite extension in central neurons: a novel role for the receptor tyrosine kinases Ror1 and Ror2. J Cell Sci 118:433–446. doi: 10.1242/jcs.01622

250. Piera-Velazquez S, Li Z, Jimenez SA (2011) Role of Endothelial-Mesenchymal Transition (EndoMT) in the Pathogenesis of Fibrotic Disorders. Am J Pathol 179:1074–1080. doi: 10.1016/j.ajpath.2011.06.001

251. Hudecek M, Schmitt TM, Baskar S, et al (2010) The B-cell tumor–associated antigen ROR1 can be targeted with T cells modified to express a ROR1-specific chimeric antigen receptor. Blood 116:4532–4541. doi: 10.1182/blood-2010-05-283309

252. Medici D, Hay ED, Olsen BR (2008) Snail and Slug Promote Epithelial-Mesenchymal Transition through β-Catenin–T-Cell Factor-4-dependent Expression of Transforming Growth Factor-β3. Mol Biol Cell 19:4875–4887. doi: 10.1091/mbc.E08-05-0506

253. Kim BJ, Hancock BM, Bermudez A, et al (2015) Bacterial induction of Snail1 contributes to blood-brain barrier disruption. J Clin Invest 125:2473–2483. doi: 10.1172/JCI74159

254. Clarke TB, Francella N, Huegel A, Weiser JN (2011) Invasive Bacterial Pathogens Exploit TLR-Mediated Downregulation of Tight Junction Components to Facilitate Translocation across the Epithelium. Cell Host Microbe 9:404–414. doi: 10.1016/j.chom.2011.04.012

255. Blumenthal A, Ehlers S, Lauber J, et al (2006) The Wingless homolog WNT5A and its receptor Frizzled-5 regulate inflammatory responses of human mononuclear cells induced by microbial stimulation. Blood 108:965–973.

256. Naskar D, Maiti G, Chakraborty A, et al (2014) Wnt5a-Rac1-NF-κB homeostatic circuitry sustains innate immune functions in macrophages. J Immunol Baltim Md 1950 192:4386–4397. doi: 10.4049/jimmunol.1302817

257. Kim J, Kim J, Kim DW, et al (2010) Wnt5a induces endothelial inflammation via beta-catenin-independent signaling. J Immunol Baltim Md 1950 185:1274–1282. doi: 10.4049/jimmunol.1000181

258. Sen M (2005) Wnt signalling in rheumatoid arthritis. Rheumatology 44:708–713. doi: 10.1093/rheumatology/keh553

259. Sen M, Lauterbach K, El-Gabalawy H, et al (2000) Expression and function of wingless and frizzled homologs in rheumatoid arthritis. Proc Natl Acad Sci U S A 97:2791–2796.

260. George SJ (2008) Wnt pathway: a new role in regulation of inflammation. Arterioscler Thromb Vasc Biol 28:400–402. doi: 10.1161/ATVBAHA.107.160952

261. Ouchi N, Higuchi A, Ohashi K, et al (2010) Sfrp5 is an anti-inflammatory adipokine that modulates metabolic dysfunction in obesity. Science 329:454–457. doi: 10.1126/science.1188280

262. Pereira C, Schaer DJ, Bachli EB, et al (2008) Wnt5A/CaMKII signaling contributes to the inflammatory response of macrophages and is a target for the antiinflammatory action of activated protein C and interleukin-10. Arterioscler Thromb Vasc Biol 28:504–510. doi: 10.1161/ATVBAHA.107.157438

263. Reischl J, Schwenke S, Beekman JM, et al (2007) Increased expression of Wnt5a in psoriatic plaques. J Invest Dermatol 127:163–169. doi: 10.1038/sj.jid.5700488

264. Möller M, Nebel A, Helden PD van, et al (2010) Analysis of eight genes modulating interferon gamma and human genetic susceptibility to tuberculosis: a case-control association study. BMC Infect Dis 10:154. doi: 10.1186/1471-2334-10-154

265. Adzhubei IA, Schmidt S, Peshkin L, et al (2010) A method and server for predicting damaging missense mutations. Nat Methods 7:248–249. doi: 10.1038/nmeth0410-248

266. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081. doi: 10.1038/nprot.2009.86

267. Baskar S, Kwong KY, Hofer T, et al (2008) Unique cell surface expression of receptor tyrosine kinase ROR1 in human B-cell chronic lymphocytic leukemia. Clin Cancer Res Off J Am Assoc Cancer Res 14:396–404. doi: 10.1158/1078-0432.CCR-07-1823

268. Glick D, Barth S, Macleod KF (2010) Autophagy: cellular and molecular mechanisms. J Pathol 221:3–12. doi: 10.1002/path.2697

269. Behrends C, Sowa ME, Gygi SP, Harper JW (2010) Network organization of the human autophagy system. Nature 466:68–76. doi: 10.1038/nature09204

270. Songane M, Kleinnijenhuis J, Alisjahbana B, et al (2012) Polymorphisms in autophagy genes and susceptibility to tuberculosis. PloS One 7:e41618. doi: 10.1371/journal.pone.0041618

271. Thwaites GE, Simmons CP, Quyen NTH, et al (2003) Pathophysiology and Prognosis in Vietnamese Adults with Tuberculous Meningitis. J Infect Dis 188:1105–1115. doi: 10.1086/378642

272. Brass AL, Huang I-C, Benita Y, et al (2009) The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. Cell 139:1243–1254. doi: 10.1016/j.cell.2009.12.017

273. Desai TM, Marin M, Chin CR, et al (2014) IFITM3 Restricts Influenza A Virus Entry by Blocking the Formation of Fusion Pores following Virus-Endosome Hemifusion. PLOS Pathog 10:e1004048. doi: 10.1371/journal.ppat.1004048

274. Feeley EM, Sims JS, John SP, et al (2011) IFITM3 Inhibits Influenza A Virus Infection by Preventing Cytosolic Entry. PLOS Pathog 7:e1002337. doi: 10.1371/journal.ppat.1002337

275. John SP, Chin CR, Perreira JM, et al (2013) The CD225 domain of IFITM3 is required for both IFITM protein association and inhibition of influenza A virus and dengue virus replication. J Virol 87:7837–7852. doi: 10.1128/JVI.00481-13

276. Ranjbar S, Haridas V, Jasenosky LD, et al (2015) A Role for IFITM Proteins in Restriction of Mycobacterium tuberculosis Infection. Cell Rep 13:874–883. doi: 10.1016/j.celrep.2015.09.048

277. Dunkelberger JR, Song W-C (2009) Complement and its role in innate and adaptive immune responses. Cell Res 20:34–50. doi: 10.1038/cr.2009.139

278. Sjöberg AP, Trouw LA, McGrath FDG, et al (2006) Regulation of Complement Activation by C-Reactive Protein: Targeting of the Inhibitory Activity of C4b-Binding Protein. J Immunol 176:7612–7620. doi: 10.4049/jimmunol.176.12.7612

279. Blom AM, Villoutreix BO, Dahlbäck B (2004) Complement inhibitor C4b-binding protein—friend or foe in the innate immune system? Mol Immunol 40:1333–1346. doi: 10.1016/j.molimm.2003.12.002

280. Ferguson JS, Weis JJ, Martin JL, Schlesinger LS (2004) Complement Protein C3 Binding to Mycobacterium tuberculosis Is Initiated by the Classical Pathway in Human Bronchoalveolar Lavage Fluid. Infect Immun 72:2564–2573. doi: 10.1128/IAI.72.5.2564-2573.2004

281. Wang C, Li Y-Y, Li X, et al (2014) Serum complement C4b, fibronectin, and prolidase are associated with the pathological changes of pulmonary tuberculosis. BMC Infect Dis 14:52. doi: 10.1186/1471-2334-14-52

282. Jarva H, Ram S, Vogel U, et al (2005) Binding of the Complement Inhibitor C4bp to Serogroup B Neisseria meningitidis. J Immunol 174:6299–6307. doi: 10.4049/jimmunol.174.10.6299

283. Davila S, Wright VJ, Khor CC, et al (2010) Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease. NatGenet 42:772–776.

284. Kobayashi K, Hernandez LD, Galán JE, et al (2002) IRAK-M Is a Negative Regulator of Toll-like Receptor Signaling. Cell 110:191–202. doi: 10.1016/S0092-8674(02)00827-9

285. Almeida AS, Lago PM, Boechat N, et al (2009) Tuberculosis is associated with a down-modulatory lung immune response that impairs Th1-type immunity. J Immunol Baltim Md 1950 183:718–731. doi: 10.4049/jimmunol.0801212

286. Pathak SK, Basu S, Bhattacharyya A, et al (2005) Mycobacterium tuberculosis lipoarabinomannan-mediated IRAK-M induction negatively regulates Toll-like receptor-dependent interleukin-12 p40 production in macrophages. J Biol Chem 280:42794–42800. doi: 10.1074/jbc.M506471200

287. Rajaram MVS, Ni B, Dodd CE, Schlesinger LS (2014) Macrophage immunoregulatory pathways in tuberculosis. Semin Immunol 26:471–485. doi: 10.1016/j.smim.2014.09.010

288. Jeyanathan M, McCormick S, Lai R, et al (2014) Pulmonary M. tuberculosis infection delays Th1 immunity via immunoadaptor DAP12-regulated IRAK-M and IL-10 expression in antigen-presenting cells. Mucosal Immunol 7:670–683. doi: 10.1038/mi.2013.86

289. LaFramboise T (2009) Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res 37:4181–4193. doi: 10.1093/nar/gkp552

290. Tian C, Gregersen PK, Seldin MF (2008) Accounting for ancestry: population substructure and genome-wide association studies. Hum Mol Genet 17:R143–R150. doi: 10.1093/hmg/ddn268

291. Skol AD, Scott LJ, Abecasis GR, Boehnke M (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet 38:209–213. doi: 10.1038/ng1706

292. Anderson CA, Pettersson FH, Clarke GM, et al (2010) Data quality control in genetic case-control association studies. Nat Protoc 5:1564–1573. doi: 10.1038/nprot.2010.116

293. Borgdorff MW, Nagelkerke NJ, Dye C, Nunn P (2000) Gender and tuberculosis: a comparison of prevalence surveys with notification data to explore sex differences in case detection. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis 4:123–132.

294. Hamid Salim MA, Declercq E, Van Deun A, Saki K a. R (2004) Gender differences in tuberculosis: a prevalence survey done in Bangladesh. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis 8:952–957.

295. Bittles AH, Black ML (2010) Consanguinity, human evolution, and complex diseases. Proc Natl Acad Sci 107:1779–1786. doi: 10.1073/pnas.0906079106

296. Weiss KM, Clark AG (2002) Linkage disequilibrium and the mapping of complex human traits. Trends Genet 18:19–24. doi: 10.1016/S0168-9525(01)02550-1

297. Price AL, Patterson NJ, Plenge RM, et al (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909.

298. Chang CC, Chow CC, Tellier LC, et al (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4:7. doi: 10.1186/s13742-015-0047-8

299. Hong EP, Park JW (2012) Sample Size and Statistical Power Calculation in Genetic Association Studies. Genomics Inform 10:117–122. doi: 10.5808/GI.2012.10.2.117

300. Klein RJ (2007) Power analysis for genome-wide association studies. BMC Genet 8:58. doi: 10.1186/1471-2156-8-58

301. Gravel S (2012) Population Genetics Models of Local Ancestry. Genetics 191:607–619. doi: 10.1534/genetics.112.139808

302. Wang X, Zhu X, Qin H, et al (2011) Adjustment for local ancestry in genetic association analysis of admixed populations. Bioinformatics 27:670–677. doi: 10.1093/bioinformatics/btq709

303. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664. doi: 10.1101/gr.094052.109

304. Quintana-Murci L, Harmant C, Quach H, et al (2010) Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. AmJHumGenet 86:611–620.

305. Patterson N, Petersen DC, van der Ross RE, et al (2010) Genetic structure of a unique admixed population: implications for medical research. Hum Mol Genet 19:411–419. doi: 10.1093/hmg/ddp505

306. Chimusa ER, Zaitlen N, Daya M, et al (2014) Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum Mol Genet 23:796–809. doi: 10.1093/hmg/ddt462

307. The 1000 Genomes Project (2007) Meeting Report: A workshop to plan a deep catalog of human genetic variation.

308. Consortium TIH 3 (2010) Integrating common and rare genetic variation in diverse human populations. Nature 467:52–58. doi: 10.1038/nature09298

309. Hoggart CJ, Parra EJ, Shriver MD, et al (2003) Control of confounding of genetic associations in stratified populations. AmJHumGenet 72:1492–1504.

310. Clarke GM, Anderson CA, Pettersson FH, et al (2011) Basic statistical analysis in genetic case-control studies. Nat Protoc 6:121–133. doi: 10.1038/nprot.2010.182

311. Lunetta KL (2008) Genetic Association Studies. Circulation 118:96–101. doi: 10.1161/CIRCULATIONAHA.107.700401

312. Karolchik D, Hinrichs AS, Furey TS, et al (2004) The UCSC Table Browser data retrieval tool. Nucleic Acids Res 32:D493–D496. doi: 10.1093/nar/gkh103

313. Nhu NTQ, Heemskerk D, Thu DDA, et al (2014) Evaluation of GeneXpert MTB/RIF for Diagnosis of Tuberculous Meningitis. J Clin Microbiol 52:226–233. doi: 10.1128/JCM.01834-13

314. Salie M, Merwe L van der, Möller M, et al (2014) Associations Between Human Leukocyte Antigen Class I Variants and the Mycobacterium tuberculosis Subtypes Causing Disease. J Infect Dis 209:216–223. doi: 10.1093/infdis/jit443

315. Mefford J, Witte JS (2012) The Covariate's Dilemma. PLOS Genet 8:e1003096. doi: 10.1371/journal.pgen.1003096

316. Segura V, Vilhjálmsson BJ, Platt A, et al (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44:825–830. doi: 10.1038/ng.2314

317. Kang HM, Sul JH, Service SK, et al (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348–354. doi: 10.1038/ng.548

318. Kang HM, Zaitlen NA, Wade CM, et al (2008) Efficient Control of Population Structure in Model Organism Association Mapping. Genetics 178:1709–1723. doi: 10.1534/genetics.107.080101

319. Manolio TA, Collins FS, Cox NJ, et al (2009) Finding the missing heritability of complex diseases. Nature 461:747–753. doi: 10.1038/nature08494

320. Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208. doi: 10.1038/ng1702

321. Zhao K, Aranzana MJ, Kim S, et al (2007) An Arabidopsis Example of Association Mapping in Structured Samples. PLOS Genet 3:e4. doi: 10.1371/journal.pgen.0030004

322. Zhang Z, Ersoz E, Lai C-Q, et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360. doi: 10.1038/ng.546

323. Guey LT, Kravic J, Melander O, et al (2011) Power in the phenotypic extremes: a simulation study of power in discovery and replication of rare variants. Genet Epidemiol 35:236–246. doi: 10.1002/gepi.20572

324. Gibson G (2012) Rare and common variants: twenty arguments. Nat Rev Genet 13:135–145. doi: 10.1038/nrg3118

325. Puzyrev VP, Freidin MB, Rudko AA, et al (2002) [Polymorphisms of the candidate genes for genetic susceptibility to tuberculosis in the Slavic population of Siberia: a pilot study]. MolBiol(Mosk) 36:788–791.

326. Deng HW, Chen WM, Recker RR (2001) Population admixture: detection by Hardy-Weinberg test and its quantitative effects on linkage-disequilibrium methods for localizing genes underlying complex traits. Genetics 157:885–897.

327. Heinze G, Schemper M (2002) A solution to the problem of separation in logistic regression. Stat Med 21:2409–2419. doi: 10.1002/sim.1047

328. Getahun H, Gunneberg C, Granich R, Nunn P (2010) HIV infection-associated tuberculosis: the epidemiology and the response. Clin Infect Dis Off Publ Infect Dis Soc Am 50 Suppl 3:S201-207. doi: 10.1086/651492

329. Sobota RS, Stein CM, Kodaman N, et al (2016) A Locus at 5q33.3 Confers Resistance to Tuberculosis in Highly Susceptible Individuals. Am J Hum Genet 98:514–524. doi: 10.1016/j.ajhg.2016.01.015

Appendices

Appendix 1

Appendix 1: Gene-sets from the SKAT-O analysis with p-values < 0.01. These gene-sets were used as input for IPA core analysis

Gene-set	Unadjusted P-value	Number of SNPs in Gene-set	Number of SNPs Tested
ZFHX3	0.000463	8	8
CYP4F2	0.001001	2	2
PHC1	0.001624	3	2
PSG5	0.002012	3	3
MUC17	0.002507	6	6
RIF1	0.002656	7	7
PIGG	0.00357	5	5
ESPL1	0.004212	5	5
ITPR3	0.004462	4	4
DCHS2	0.004487	3	3
C2orf61	0.004728	1	1
LHFPL5	0.005362	1	1
MRO	0.005362	1	1
SERGEF	0.005362	1	1
TOPBP1	0.005781	2	2
FARP2	0.006122	3	3
CHL1	0.006647	4	4
USP6	0.006652	5	5
ADCK4	0.006676	4	4
NOD2	0.006711	3	3
RFX8	0.007084	4	4
LTBP4	0.007129	7	7
NPIPB3	0.00726	3	3
CTAGE4	0.007659	14	11
RPGRIP1L	0.008034	5	5
KIDINS220	0.00829	1	1
CC2D1A	0.008785	4	4
ZNF107	0.008786	1	1
ALDH3B1	0.00887	2	2
KIAA1919	0.008925	2	2
OR4D9	0.009327	1	1
FSTL5	0.009362	3	3
ESRRA	0.009365	7	7

KCP	0.009549	13	13
COL11A1	0.009604	5	5
OR1J2	0.009747	2	2
FRMD5	0.009894	1	1
TMEM150B	0.009907	1	1
GCKR	0.009942	1	1

Appendix 2

Appendix 2: Gene-sets from the SKAT Common Rare analysis with p-values < 0.01. These gene-sets were used as input for IPA core analysis

Gene-set	Unadjusted P-value	Number of SNPs in Gene-set	Number of SNPs Tested	Number of SNPs, MAF < 0.01 (Rare)	Number of SNPs, MAF > 0.01 (Common)
CCP110	5.89E-06	3	3	1	2
OR2A5	1.49E-05	5	5	2	3
CCNI2	0.000115	2	2	1	1
APOBEC4	0.000146	5	5	0	5
C4BPA	0.000254	6	6	5	1
PLBD2	0.000272	3	3	2	1
DOCK6	0.000274	7	7	6	1
ZNF804B	0.000369	9	9	8	1
SERGEF	0.000369	3	3	3	0
LDLRAD2	0.000422	1	1	0	1
OR2T5	0.000607	11	11	3	8
APOF	0.000632	2	2	1	1
ROR1	0.00066	1	1	0	1
SLC45A2	0.000789	3	3	2	1
CFAP44	0.000826	18	18	17	1
ABCA6	0.000831	10	10	6	4
ATG10	0.000856	3	3	0	3
PAN2	0.000901	2	2	1	1
ZFYVE16	0.001004	8	8	5	3
CCDC18	0.001088	5	5	3	2
PNPT1	0.001122	5	5	3	2
PROB1	0.001234	7	7	6	1
HAUS6	0.001368	8	8	7	1
TMED8	0.001409	3	3	1	2
CKAP2L	0.001452	7	7	6	1
RSPO2	0.001503	1	1	0	1

RRP36	0.001523	1	1	0	1
MARCO	0.001755	2	2	1	1
FAM208A	0.001793	5	5	3	2
MTAP	0.001892	1	1	0	1
NPHP4	0.002075	10	10	8	2
GDF5	0.002125	3	3	2	1
EXD3	0.002176	19	19	16	3
SMUG1	0.002186	2	2	2	0
G6PC	0.002384	1	1	1	0
CTAGE15	0.002459	7	6	3	3
SPAG16	0.002536	6	6	4	2
OXSR1	0.002593	1	1	0	1
TNNI1	0.002677	1	1	1	0
FAM135A	0.002685	3	3	2	1
SLC25A39	0.002694	1	1	0	1
TMEM19	0.002738	1	1	1	0
ASB17	0.002918	1	1	0	1
GUCA1C	0.002933	4	4	2	2
IFNK	0.003147	2	2	1	1
CYFIP1	0.003199	2	2	1	1
KRT25	0.003233	2	2	1	1
ARHGEF17	0.003428	8	8	7	1
DPY19L1	0.003455	1	1	0	1
KLK10	0.00357	2	2	1	1
CLC	0.003578	1	1	0	1
CNTRL	0.003599	8	8	6	2
ANLN	0.003646	4	4	3	1
KRT20	0.003692	2	2	1	1
NAT9	0.003743	2	2	1	1
EVPLL	0.003795	3	3	2	1
TEX15	0.003905	17	17	12	5
MYO7A	0.003988	9	9	6	3
KHDC3L	0.004007	2	2	1	1
MACF1	0.00429	14	14	12	2
SRRD	0.004298	2	2	1	1
VSIG10	0.004348	3	3	2	1
SMCR8	0.0044	6	6	4	2
TMCO4	0.00442	7	7	5	2
OR6C74	0.004431	7	7	3	4
CLSPN	0.004446	9	9	8	1
DDHD2	0.004602	1	1	1	0

B3GNT2	0.004623	1	1	1	0
GNAT2	0.004625	1	1	1	0
FAM135B	0.004678	4	4	2	2
R3HDM2	0.004769	1	1	1	0
BBS9	0.004783	7	7	6	1
AK4	0.004789	3	3	2	1
NMB	0.004869	1	1	1	0
DDX11	0.004949	10	10	7	3
SCYL3	0.005048	2	2	1	1
CYBRD1	0.005057	2	2	1	1
GPR56	0.005066	3	3	2	1
OR7G3	0.005156	9	9	7	2
SAMD15	0.005164	9	9	6	3
QSOX2	0.005205	2	2	2	0
MAP2	0.005219	4	4	3	1
HEXA	0.005255	2	2	1	1
GNRH1	0.005362	1	1	1	0
ZNF133	0.005362	2	2	2	0
PLA2G4D	0.005363	10	10	8	2
GSDMA	0.005428	5	5	1	4
PATE1	0.005447	2	2	1	1
FCRL3	0.005501	4	4	3	1
CBY3	0.005623	3	3	2	1
TMEM211	0.005643	2	2	2	0
OR6Y1	0.005709	5	5	4	1
KLK2	0.005909	2	2	1	1
CRNKL1	0.005945	6	6	3	3
KIDINS220	0.005971	2	2	2	0
ADCK5	0.005977	1	1	1	0
C1orf94	0.006168	5	5	2	3
DDX52	0.006202	2	2	1	1
EXO5	0.006396	3	3	2	1
PCDHA9	0.006427	13	13	7	6
DDX43	0.006446	4	4	2	2
HHATL	0.006594	4	4	3	1
WSB2	0.006646	1	1	1	0
FRMD5	0.006747	1	1	1	0
G6PC3	0.006747	1	1	1	0
PACS2	0.006751	5	5	4	1
LHFPL5	0.006842	1	1	1	0
NR5A1	0.006862	1	1	0	1

CCDC137	0.006928	8	8	5	3
AGMO	0.007152	5	5	4	1
ZFP62	0.007167	2	2	1	1
PPY	0.007178	1	1	1	0
GRIP2	0.007196	13	12	7	5
CTAGE9	0.00722	11	11	9	2
HAVCR1	0.007383	5	5	4	1
TRDMT1	0.007401	2	2	1	1
NLGN4X	0.007411	1	1	0	1
PDCD1LG2	0.007552	4	3	2	1
ZNF267	0.007637	1	1	0	1
PPIAL4G	0.007691	9	6	3	3
TRPM2	0.007722	12	11	10	1
HOXC4	0.007775	1	1	0	1
FAM71F2	0.008107	5	5	2	3
OR2A25	0.008131	5	5	3	2
APBB2	0.008149	2	2	1	1
HDHD1	0.008232	2	2	1	1
IRAK3	0.008238	4	4	3	1
ZNF816	0.0084	5	5	4	1
MRPL37	0.008511	2	2	1	1
TAS2R9	0.008528	4	4	3	1
AMER1	0.008563	1	1	0	1
KIAA1462	0.008571	6	6	4	2
A2M	0.008629	10	9	8	1
TUBG2	0.008754	2	2	1	1
OR4K17	0.008912	3	3	2	1
CELSR2	0.008933	8	8	7	1
PDCD4	0.009141	1	1	0	1
IFITM2	0.009234	3	3	1	2
C5orf42	0.009251	8	8	7	1
C18orf21	0.009288	2	2	1	1
GPR98	0.009304	45	44	36	8
OR6C1	0.009453	6	6	4	2
GSTA2	0.009478	3	3	1	2
TMEM214	0.009542	3	3	2	1
CX3CR1	0.009569	4	4	3	1
UAP1L1	0.009624	5	5	3	2
SLC36A3	0.009629	4	4	3	1
ARHGAP9	0.009636	5	5	4	1
AGXT2	0.009701	6	6	3	3

DNMT1	0.00972	1	1	1	0

Appendix 3

Appendix 3: Principal component analysis plots with TBM cases and controls genotyped on the MEGA array compared to reference samples from the representative populations of the 1000 genomes project.